JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

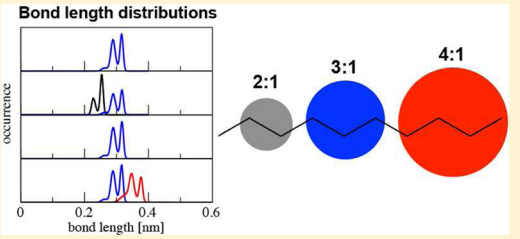
Supra-Atomic Coarse-Grained GROMOS Force Field for Aliphatic Hydrocarbons in the Liquid Phase

Andreas P. Eichenberger,[†] Wei Huang,[†] Sereina Riniker, and Wilfred F. van Gunsteren*

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, 8093 Zürich, Zürich, Switzerland

Supporting Information

ABSTRACT: A supra-atomic coarse-grained (CG) force field for liquid *n*-alkanes is presented. The model was calibrated using experimental thermodynamic data and structural as well as energetic properties for 14 *n*-alkanes as obtained from atomistic fine-grained (FG) simulations of the corresponding hydrocarbons using the GROMOS 45A3 biomolecular force field. A variation of the nonbonded force-field parameters obtained from mapping the FG interactions onto the CG degrees of freedom to fit the density and heat of vaporization to experimental values turned out to be mandatory for a correct reproduction of these data by the CG model, while the bonded force-field parameters for the CG model could be obtained from a Boltzmann-weighted fit with some variations with respect to the corresponding properties from the FG simulations mapped onto the CG degrees of freedom. The model presents 6 different CG bead types, for bead sizes from 2 to 4 distinguishing between terminal and nonterminal beads within an alkane chain (end or middle). It contains different nonbonded Lennard-Jones parameters for the interaction of CG alkanes with CG water. The CG alkane model was further tested by comparing predictions of the excess free energy, the self-diffusion constant, surface tension, isothermal compressibility, heat capacity, thermal expansion coefficient, and shear viscosity for *n*-alkanes to experimental values. The CG model offers a thermodynamically calibrated basis for the development of CG models of lipids.



INTRODUCTION

Since the first simulation of protein motion[1] simulation of the dynamics in biomolecular systems, proteins, nucleotides, carbohydrates, lipids, and the like in aqueous solution has become a standard method of investigation over the past decades.[2,3] Because an accurate quantum-mechanical treatment of the many nuclear and electronic degrees of freedom in such systems in simulations for a sufficiently long time period to observe functionally relevant conformational changes is as yet impossible using currently available computers, biomolecular systems are still generally simulated at the atomic level of resolution using classical-mechanical equations of motion and an empirical potential energy function that yields the potential energy of the system in terms of atomic coordinates.[4] Such a potential energy function is shortly called force field, and a number of general atomic-level interaction functions for biomolecules have been developed over the past decades, e.g. AMBER,[5,6] CHARMM,[7,8] and GROMOS.[9,10]

When simulating very large systems such as multiprotein complexes or membranes, simulation at the atomic level of resolution can only reach the nanosecond time scale which is insufficient to properly sample longer time motions in such systems. Therefore, much effort has more recently been spent on the development of more coarse-grained, supra-atomic level of resolution models for biomolecular systems, see e.g. refs 11−16 for reviews. Because coarse-graining involves the elimination of particular degrees of freedom from a model, some properties of the molecular system may get lost in the coarse-graining process. This implies that the choice of degrees of freedom that are to be kept in the more coarse-grained model will depend on the particular properties of the molecular system that are the focus of the investigation using computer simulation.

When can particular degrees of freedom be eliminated? They must be nonessential for the process or property of interest and must be large in number in order to have the computational gain offset the inevitable loss of accuracy through coarse-graining. Interactions governing eliminated degrees of freedom should be largely decoupled from those governing the degrees of freedom remaining after coarse-graining, leading to decoupled motions between these two sets of degrees of freedom, and the effective interactions governing the remaining degrees of freedom should be simply representable in the form of a fast computable potential energy function.[16,17] For example, the motions of aliphatic hydrogen atoms in CH, CH_2, and CH_3 groups are irrelevant for most processes of biochemical interest. By treating these groups as single united atoms,[18] the number of interaction sites is reduced by a factor 2 to 4. For alkanes and lipids this leads to almost a factor of 10 fewer pairwise nonbonded interactions at the cost of losing the dipolar interactions of CH_n-moieties and the van der Waals interactions of the hydrogen atoms. The intragroup motions of CH_n moieties are largely decoupled from other motions in such molecules, and the torsional interactions involving these hydrogen atoms can be effectively included in the torsional

Received: March 30, 2015

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

interaction function around the same bond that only involves non-hydrogen atoms. Thus, all four conditions for proper coarse-graining are basically fulfilled in this case. For this reason the GROMOS atomic-level resolution force field represents aliphatic CH_n-groups as united atoms.

What molecular properties or processes are important for the process of interest and should thus be maintained while coarse-graining?

1. Molecular structure of the solute, e.g. the protein, and structure of the solvent or liquid.

2. Thermodynamic properties that carry volume and energetic information, such as density, heat of vaporization, and excess free energy. Thermodynamic properties that characterize a response to a change in thermodynamic state point, such as compressibility and heat capacity, are of less importance.

3. Dielectric properties, such as the static dielectric permittivity $\varepsilon(0)$ that governs the screening of Coulomb interactions.

4. Dynamic properties such as diffusion, viscosity, and molecular relaxation times are less important, because most biomolecular processes are thermodynamically, not dynamically, driven, at least when considering the time and space scales accessible to simulations.

In any case, the process of coarse-graining is likely to reduce the usefulness of the model in different ways: (*i*) The range of thermodynamic state points at which the model may be applied is generally reduced; (*ii*) The transferability of model parameters between similar but not identical moieties or compounds is usually reduced; (*iii*) The accuracy of the reproduction of measured values for properties may be reduced; (*iv*) The physical basis of a particular property or process may be changed, leading to an unphysical mechanism of the process in the coarse-grained model, an example being the use of implicit solvation models[19] in which solvent degrees of freedom are omitted and their effect represented as a function of the solute coordinates only; (*v*) The reduction of entropy and energy in the system may lead to an unphysical balance between these two quantities in the coarse-grained model.[20] In short, the combined loss of usefulness on these five counts must be made up for by a much increased computational efficiency of the coarse-grained model compared to the fine-grained one. This trade-off between atomic (fine-grained) and supra-atomic (coarse-grained) levels of resolution is negative for biomolecules such as proteins, nucleotides, and sugars because of their heterogeneous composition of atoms and polar versus nonpolar moieties, which implies a rather large loss of specificity when coarse-graining from the atomic to a supra-atomic level of resolution in combination with a rather modest gain in computational efficiency. This trade-off is positive when coarse-graining solvent degrees of freedom from the atomic level to a supra-molecular level of resolution.[21,22] Simulating proteins in water, the use of a supra-molecular coarse-grained water model leads to an order of magnitude more efficient simulations while maintaining structural and energetic properties of the protein.[23,24] When coarse-graining models for lipids, the trade-off may also be positive because of their abundance in membranes.

Different supra-atomic coarse-grained models for lipids are available.[25−27] The models by Marrink et al.[25] use spherical beads in conjunction with a Lennard-Jones model for coarse-grained water. The models by Essex and co-workers[26] use ellipsoidal beads in conjugation with a soft sticky coarse-grained water model, and the model by Jakobsson and co-workers[27] uses a Morse potential for the nonbonded potential energy. These models are not compatible with the atomic-level GROMOS force field which we would like to use in multigraining simulations of atomic-level proteins in membranes composed of supra-atomic lipids embedded in supra-molecular water. In order to develop a supra-atomic coarse-grained force field for lipids that can be made compatible with the atomic-level GROMOS force field for biomolecules, we investigated some structural and thermodynamic properties of the supra-atomic MARTINI force field for biomolecules[28] using a range of alkanes in the liquid phase,[29] which served as representation of the properties of the aliphatic tails of lipids. Experimental properties such as density and heat of vaporization were not well reproduced, and bead−bead radial distribution functions appeared to be overstructured compared to the corresponding functions derived from atomic-level simulations of liquid alkanes. It was concluded that these discrepancies could be resolved by a reparametrization of the supra-atomic force field for aliphatic chains.[29] The results of this reparametrization are reported here.

The strategy for the development of the GROMOS force fields has been described in ref 11: After having chosen a functional form for the various terms in the potential energy function, the force-field parameter values have to be determined. For the bond-length and bond-angle terms this can be done using X-ray diffraction and spectroscopic data on small molecules. The parameters of the torsional-angle interaction term and partial atomic charges can be obtained from quantum-chemical calculations on small model compounds in the gas phase. However, since accurate quantum-chemical calculations for molecules in the liquid phase are not yet possible, the nonbonded interaction parameters, partial charges q_i and van der Waals parameters $C_{12}(i, j)$ and $C_6(i, j)$ for atoms or beads i and j, have to be further calibrated using thermodynamic data such as density, heat of vaporization, excess free energy, and dielectric data measured for liquids of the small molecules used for parameter calibration.

This strategy that was applied to develop the atomic-level GROMOS force fields[9,10] can also be applied to develop supra-atomic or supra-molecular force fields.[21,22,30] However, in the case of supra-molecular models, some properties such as the excess free energy of the liquid cannot be used for parametrization, and some technical issues due to the supra-molecular character of the coarse-grained beads emerge.[21] In the case of supra-atomic models, e.g. for lipids, the difficulty lies in the scarcity of structural, thermodynamic, and dielectric experimental data on such molecules in the fluid phase. The surface area of bilayers of lipids in water and the C−H order parameters of CH_2−moieties in the aliphatic tails as derived from NMR experiments are often used to test the quality of membrane simulations and the force field used,[31] but these data do not possess a high accuracy[31] and, due to their low number compared to the number of degrees of freedom of a lipid, do not contribute data that can be used for force-field parameter calibration. For liquid alkanes there is thermodynamic experimental data which can be used in the parameter calibration. Since structural experimental data, apart from trans/gauche ratios derived from NMR experiments,[32] is lacking for these flexible molecules, one can resort to the use of structural data as generated in simulations using an atomic-level model and force field when calibrating parameters of a supra-atomic model and force field for the same compound.

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

The caveat is here that the inaccuracies of the atomic-level force field may get propagated to the supra-atomic level one.

Based on these considerations, the strategy followed here to obtain a supra-atomic GROMOS force field for alkanes in the liquid phase is the following one:

1. Molecular dynamics (MD) simulations of 14 *n*-alkanes in the liquid phase at ambient temperature and pressure are carried out using the atomic-level of resolution GROMOS force field 45A3[33] in order to obtain fine-grained configurational ensembles.

2. If these atomic-level configurational ensembles reproduce the experimental densities and heat of vaporization of these 14 liquids, they can be used as calibration data for structural properties at the supra-atomic, coarse-grained level of modeling. To this end the atomic-level configurations are to be mapped onto supra-atomic configurations. For example, two, three, or four atoms are mapped onto or represented by their center of geometry or their center of mass. Using such an atom-to-bead mapping the structural, energetic, dielectric, and dynamic properties of the atomic-level MD trajectories are represented in terms of the corresponding quantities for beads representing 2 to 4 atoms. In the present case of alkanes these quantities are e.g. distance-, angle-, and torsional-angle distributions involving 2 to 4 beads, bead−bead radial distribution functions, and nonbonded interaction energies.

3. Based on these atomic-level distributions of various quantities the form of the supra-atomic level model can be chosen: How many atoms are mapped onto one bead? Are beads in the middle of the aliphatic chain to be distinguished from those terminating a chain? For various reasons[16] the functional form of the supra-atomic potential energy function that depends on the coordinates of the beads is the same as that of the atomic-level GROMOS force field. A first estimate for supra-atomic force-field parameters can be obtained from this analysis of the atomic-level MD trajectories.

4. Molecular dynamics simulations of *n*-alkanes in the liquid phase at ambient temperature and pressure are carried out using the supra-atomic level model and estimated force-field parameters.

5. If these coarse-grained configurational ensembles do not reproduce the structural and thermodynamic target data, e.g. the density, the heat of vaporization, and the free enthalpy of hydration, a further calibration of the supra-atomic force-field parameters is required. For the *n*-alkanes, the ε and σ Lennard-Jones parameters or the C_{12} and C_6 van der Waals parameters had to be varied to reproduce the densities and heats of vaporization, and separate C_{12} parameters for the interaction with CG water had to be determined.

6. If the target properties are reproduced satisfactorily by one of the coarse-grained models, this model and parameter set is further tested by the calculation and comparison to experimental values of quantities that were not included in the set of target data.

In the current case of the development of a coarse-grained supra-atomic GROMOS force field for *n*-alkanes in the liquid phase, six bead types were defined, representing 2, 3, or 4 united atoms and distinguishing between beads within the chain having two bound neighbor beads and beads terminating a chain having one bound neighbor bead. The use of a combination of different atom-to-bead mappings in one chain allows for a simple comparison of fine-grained and coarse-grained configurations for arbitrary chain lengths.

METHODS

All fine-grained (FG) and coarse-grained (CG) molecular dynamics (MD) simulations as well as the analysis have been performed using the GROMOS software package for biomolecular simulations.[34−36] Some analysis methods concerning the calculations of CG properties calculated from mapped FG alkane configurations, such as the calculation of nonbonded bead−bead energies and the computation of radial distribution functions, needed some additional GROMOS++ code not included in the officially released GROMOS version.[37]

Atomistic FG Model. Interactions of the 14 FG alkanes (butane to heptadecane) were modeled using the GROMOS 45A3 force-field parameter set.[33] This model consists of five atom types, CH_n with n = 0, 1, ..., 4, ready to be used for the simulation of aliphatic hydrocarbons in the condensed phase. All of them have zero partial charge and are nonpolarizable. In the current work we only used CH_2 and CH_3 FG atoms.

Generally, at the FG atomic level of resolution, the bond lengths of a biomolecular solute are kept rigid in order to allow for larger MD time steps, 2 fs instead of 0.5 fs. If this is not the case, a quartic bond-stretching potential energy term V^b is used to calculate forces between covalently bound neighbor atoms[38]

$$V^{(b)}(b; K_b, b_0) = \frac{1}{4}K_b(b^2 - b_0^2)^2 \tag{1}$$

where $b = |r_i - r_j| = r_{ij}$ is the distance between atoms i and j, and the two force-field parameters K_b and b_0 indicate the force constant and the reference bond length. The potential energy term for bond angles is defined as[38]

$$V^{(a)}(\theta; K_\theta, \theta_0) = \frac{1}{2}K_\theta(\cos\theta - \cos\theta_0)^2 \tag{2}$$

with the angle-force constant K_θ and the reference bond angle θ_0. GROMOS also provides the use of harmonic bond-stretching and angle-harmonic bond-angle terms. However, we stick to the quartic and cosine-harmonic form, which is still the standard choice in GROMOS. In addition, there is a proper dihedral angle torsional force-field term[38] to mimic the rotational barriers along a bond

$$V^{(t)}(\varphi; K_\varphi, \delta, m) = K_\varphi(1 + \cos\delta\cos(m\varphi)) \tag{3}$$

with $\delta = 0, \pi$, K_φ the force constant, φ the angle of four neighboring, covalently bound atoms, and m the multiplicity. The 45A3 force field[33] also offers an improper dihedral-angle potential energy term[38] used to preserve the stereochemistry of branched alkanes.

The nonbonded van der Waals interaction between two particles i and j is modeled using a Lennard-Jones (12/6) functional form

$$V_{i,j}^{(vdw)}(r_{i,j}) = \left(\frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^{6}}\right) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{4}$$

where the Lennard-Jones (LJ) parameters C_{12} (i,j) and C_6 (i,j) of atom type i and j are combined using the geometric combination rule

$$C_{12}(i,j) = \sqrt{C_{12}(i,i)}\sqrt{C_{12}(j,j)} \text{ and } C_6(i,j) = \sqrt{C_6(i,i)}\sqrt{C_6(j,j)} \tag{5}$$

Covalently bound first- and second-neighbor atoms are excluded from the nonbonded interaction since their behavior

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Number of Molecules and Atoms in the FG Simulations of *n*-Alkanes[c]

	number of		ρ (kg m^{-3})		ΔH_{vap} (kJ mol^{-1})	
alkane	molecules	atoms	exp	calc	exp	calc
butane	2500	10000	573.0[a]	577.8	21.62	21.20
pentane	2000	10000	626.2[b]	621.1	26.43	26.35
hexane	1667	10002	660.6	654.7	31.55	31.71
heptane	1429	10003	679.5	678.7	36.55	36.67
octane	1250	10000	698.6	698.0	41.49	41.38
nonane	1112	10008	719.2[b]	712.6	46.44	46.53
decane	1000	10000	726.6	724.5	51.37	51.56
undecane	910	10010	740.2[b]	734.6	56.33	56.29
dodecane	834	10008	749.5[b]	743.2	61.29	61.53
tridecane	770	10010	756.4[b]	750.4	66.23	66.15
tetradecane	715	10010	759.6[b]	756.4	71.17	71.60
pentadecane	667	10005	768.5[b]	762.0	76.15	76.20
hexadecane	626	10016	770.1	766.8	81.09	81.16
heptadecane	589	10013	778.0[b]	771.4	86.19	86.03

[a]Experimental density measured at a pressure >1 atm. [b]Experimental density measured at a temperature of 293 K. [c]Density ρ and heat of vaporization ΔH_{vap} for liquid alkanes at 298 K and 1 atm, unless indicated otherwise. Experimental values were taken from ref 51.

is described by and would conflict with the bonded interaction terms. Third-neighbors do interact according to eq 4 but use scaled LJ interaction parameters, $CS_{12}(i, i)$ and $CS_6(i, i)$.

Atomistic FG Simulations. If the initial configurations of the *n*-alkanes would not yield the experimentally derived averaged trans/gauche ratio, a rather long MD equilibration period would be required to obtain the proper trans/gauche ratio in the liquid phase. To reduce the equilibration time, the initial configurations of individual molecules were taken from the last N_f molecule configurations 20 ps apart in a 100 ns stochastic dynamics (SD) simulation trajectory of a single alkane molecule at 298.15 K using a friction coefficient γ = 50 ps^{-1}. N_f corresponds to the number of molecules as indicated in Table 1 contained in the periodic computational box. They were randomly distributed in cubic boxes using the GROMOS++ program ran_box.[36] The individual box-edge lengths were chosen between 7.5 nm (butane) and 6.7 nm (heptadecane) to yield the experimental densities as indicated in Table 1. The molecular configurations were then energy minimized under minimum image periodic boundary conditions to relieve high energy contacts between individual molecules. Initial velocities were assigned from a Maxwell−Boltzmann distribution at 1 K, and the temperature was continuously raised to 298.15 K during the first 500 ps of equilibration time at constant volume, followed by another 500 ps simulation at constant volume and constant temperature (298.15 K). Another 2 ns simulation at constant temperature and pressure (298.15 K, 1 atm) was appended to complete the equilibration of the different alkane boxes, before continuing the simulations under identical conditions for another 10 ns saving the atom configurations and energies every 5 ps for analysis.

The weak-coupling algorithm[39] was used for constant temperature and pressure simulations with corresponding coupling times of τ_T = 0.1 ps and τ_p = 0.5 ps, respectively, and an isothermal compressibility ranging from 0.00134 to 0.00523 (kJ mol^{-1} nm^{-3})$^{-1}$ corresponding to experimental values. All bond lengths were kept constant using the SHAKE algorithm[40] with a relative geometric tolerance of 10^{-4}, allowing for an integration time step of 2 fs when solving the equations of motion using the leapfrog algorithm.[41] Nonbonded (van der Waals) interactions were handled adopting a triple-range scheme with cutoff radii of 0.8/1.4 nm: interactions within the short-range cutoff were evaluated every time step, and the intermediate range interactions were updated every fifth time step of 2 fs. In the simulations of an alkane molecule solvated in SPC water, the long-range electrostatic interactions were approximated by a reaction-field force[42] representing a continuum with a relative dielectric permittivity ε_{RF} of[43] 61. The center of mass translation and rotation were removed every 2 ps to avoid a flying ice cube.[44]

Analysis of FG Configurations in Terms of CG Beads. Individual values for the bonded and nonbonded CG force-field parameters were obtained from an analysis of the atomic-level alkane simulation trajectories mapped onto CG beads. To this end, the atoms of the FG alkane molecules were mapped to CG beads, each bead replacing 2 to 4 FG atoms. The position of the CG bead was set to the center of mass of the replaced FG atoms

$$\mathbf{r}_{CG} = \frac{1}{M}\sum_{i=1}^{N} m_i \mathbf{r}_i \tag{6}$$

where N is the number of atoms to be replaced, $M = \Sigma_{i=1}^{N} m_i$ is the total mass of all replaced atoms, and $\mathbf{r}_i$ is the position of atom i. This mapping is not restricted to a fixed number of atoms per bead within one molecule chain. Bonded force-field parameters were obtained applying a Boltzmann weighted fitting

$$p(Q) \propto \exp\left(-\frac{V_Q}{k_B T}\right) \tag{7}$$

the quantity Q being the bond distance, the bond angle, or the torsional angle of neighboring beads, k_B being the Boltzmann constant, T = 298.15 K being the temperature, and V_Q being the corresponding potential energy as described in eqs 1−3. The reference bond length b_0^{CG} at which the quartic bond stretching potential energy is zero (compare eq 1) was chosen (Table 2) to correspond to the average CG bond length as calculated from the mapped FG simulation trajectories and was increased by a factor of 1.0134 in order to obtain the correct average CG bond lengths in simulations without CG bond-length constraints.

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Bonded Force-Field Parameters for Coarse-Grained *n*-Alkanes[a]

bond stretching	b_0 [nm]		K_b [kJ mol^{-1} nm^{-4}]
C2−C2	0.248		100000
C2−C3	0.303		60000
C3−C3	0.358		25000
C3−C4	0.410		18000
C4−C4	0.463		10000
bond-angle bending	θ_0 [°]		K_θ [kJ mol^{-1}]
C2−C2−C2	150.1		53.3
C2−C3−C2	157.1		58.9
C2−C3−C3			
C3−C3−C3			
C4−C3−C4	153.8		48.1
C4−C4−C4			
C3−C2−C3	147.2		53.3
C2−C2−C3			
dihedral-angle torsion	cos δ	m	K_φ [kJ mol^{-1}]
C2−C2−C2−C2	−1.0	1	0.68
C2−C3−C3−C2	−1.0	1	
C2−C3−C3−C3	−1.0	1	
C3−C2−C3−C3	−1.0	1	
C3−C3−C3−C3	−1.0	1	0.74
C4−C4−C4−C4	−1.0	1	1.20

[a]Note that there is no distinction between middle beads or particles (C2M, C3M, C4M) and end beads or particles (C2E, C3E, C4E). The parameters are defined in eqs 1−3.

The force-field parameters of the nonbonded LJ interaction for two beads I and J, $C_{12}(I,J)$ and $C_6(I,J)$, were estimated from the mapped LJ interaction potential energy according to

$$V_{IJ}^{(vdw)}(r_{IJ}) = \left\langle \sum_{i=1}^{N}\sum_{j=1}^{N} V_{ij}^{(vdw)}(r_{ij}) \right\rangle \quad (8)$$

where N is the number of atoms per bead, $V_{ij}^{(vdw)}(r_{ij})$ is as described in eq 4, and angular brackets indicate the average over all intermolecular bead−bead pairs. They are shown in Table S1.

Supra-Atomic CG Model. Since the CG model for linear alkanes is an intermediate step on the way toward a CG model for lipid bilayers to be used in biomolecular MD simulations, it is required to be compatible with existing and upcoming GROMOS force fields at the FG level for solutes as well as at the FG and CG levels for solvents. Therefore, the functional forms of the bonded and nonbonded potential energy terms used in the FG model, eqs 1 to 5, are also chosen for the CG alkane model but applied to bonded and nonbonded interactions of beads instead of atoms. In the CG alkane model, not only first and second neighbors are excluded from the summation in calculating nonbonded interactions but also the third neighbors are excluded. A CG bead was chosen to represent 2 to 4 FG atoms yielding 6 CG atom types: C2M, C2E, C3M, C3E, C4M, and C4E where the indices E (end) and M (middle) indicate the position of the corresponding bead within the alkane chain.

Supra-Atomic CG Simulations. Initial configurations of the first CG alkane simulations, i.e. using the bonded force-field parameters and LJ parameters directly calculated from the mapped FG alkane simulation trajectories by considering intermolecular geometries and intermolecular interactions respectively, were generated by randomly distributing copies of a precursor molecule in a cubic box. The total number of CG molecules was identical to the number of molecules in the FG simulations, compare Table 1, while the energy minimized configuration of an FG alkane molecule having the same number of atoms as the CG alkane has number of beads, served as the precursor molecule, e.g. butane was serving as the precursor molecule of CG dodecane with bead size 3. Box sizes were chosen to yield the experimental density of the corresponding alkane (Table 1). A steepest descent energy minimization was performed applying minimum image periodic boundary conditions to resolve unfavorable intermolecular bead−bead interactions.

Initial bead velocities were assigned from a Maxwell−Boltzmann distribution at 1 K and continuously raised to 298.15 K during the first 1 ns of the simulation at constant volume. Another 1 ns simulation at constant volume and temperature (298.15 K) was followed by a simulation at constant pressure and temperature (1 atm, 298.15 K) to complete the 3 ns of equilibration. The simulation was then continued for another 6 ns, and the bead positions and velocities as well as the system energies were saved every 5 ps for analysis.

Constant temperature and pressure were maintained using the weak-coupling algorithm[39] with coupling times τ_T = 0.1 ps and τ_p = 0.5 ps, respectively, and an isothermal compressibility of 4.575 × 10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$ for all alkanes, corresponding to an averaged value for biomolecular systems. A time step of 2 fs was used when solving Newton's equation of motion using the leapfrog algorithm.[41] A much bigger time step would be possible for coarse-grained alkanes as shown in the literature.[21,28] However, we kept 2 fs since the CG model is meant for use in multigrained simulations.[24,30,45] Nonbonded interactions were handled using a single-range cutoff radius of 2 nm as it was applied in the GROMOS model of CG, polarizable solvent[21,22,30] and in mixed-grain simulations.[24,30,45] The dielectric permittivity within the cutoff sphere was set[24,30,45] to ε_{cs} = 2.5 and the reaction-field value to ε_{RF} = 78.5 in the simulations involving water. The center of mass translation and rotation were removed every 2 ps to avoid a flying ice cube.[44]

The simulation with the nonbonded parameters best reproducing the experimental density and heat of vaporization was further prolonged for another 10 ns, and bead configurations and system energies were saved every 5 ps for analysis.

Parametrization Strategy for the LJ Interactions. The LJ interactions were parametrized by altering the initial ε and σ values of the LJ interaction potential energy, eq 4, which had been obtained from the mapped FG simulation trajectories, by simply scaling these values with factors f_ε and f_σ, respectively, i.e. $\varepsilon_{new} = f_\varepsilon \cdot \varepsilon_{init}$ (Tables S2−S29).

The CXE (X = 2, 3, 4) CG particles were first parametrized using CG alkanes consisting of two such particles only, i.e. CG butane for bead type C2E, hexane for bead type C3E, and octane for bead type C4E. The scaling factors f_ε and f_σ for CXE beads are listed in Table S30.

Longer alkane chains were used to parametrize the CXM (X = 2, 3, 4) beads while keeping the ε and σ values of the end-beads constant. For each bead size, different scaling factors f_ε and f_σ were obtained for CXM from different alkanes. They were subsequently averaged according to different weights for different alkanes. The individual scaling factors f_ε and f_σ for different alkanes, the weights of each alkane, and the averaged scaling factors f_ε and f_σ are listed in Table S31.

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Butane has a boiling temperature lower than room temperature, which made it less suitable for a proper parametrization of C2 particles. CG alkanes consisting of beads of different bead sizes within one molecule were not used for the parametrization of the nonbonded interactions.

The values for the LJ parameters for the interaction between CG alkane beads and the particles of the CG water model[21] were obtained by fitting the free enthalpy of hydration ΔG_{hyd} to experimental values for the set of alkanes. Only the repulsive C_{12} parameters were scaled,[24,30,45] the parametrization strategy is similar to the one described above. First, the scaling factors of C_{12} parameter of CXE (X = 2, 3, 4) beads were determined (Table S32 in the Supporting Information), second the scaling factors of the C_{12} parameters of CXM (X = 2, 3, 4) beads were derived for every single-bead-size CG alkane separately (Table S33 in the Supporting Information), and third, the scaling factors of the C_{12} parameters of CXM (X = 2, 3, 4) beads were averaged at the end (Table S33 in the Supporting Information).

Analysis. *1. Radial Distribution Function.* The radial distribution functions g_{ee}, g_{em}, and g_{mm} were calculated from the mapped FG alkane simulation trajectories and from the CG simulation trajectories of the final alkane model according to

$$g_{\alpha\beta}(r) = \left\langle \frac{1}{4\pi\rho r^2 \Delta r (N_\beta - \delta_{\alpha\beta})} \sum_{\alpha=1}^{N_\alpha - \delta_{\alpha\beta}} \sum_{\beta=1}^{N_\beta} \delta(r_{\alpha\beta} - r) \right\rangle \tag{9}$$

where N_i is the number of beads of type $i = [\alpha, \beta]$, $r_{\alpha\beta}$ is the distance between two beads, ρ_β is the particle density of beads of type β, and Δr is the grid spacing. Angular brackets indicate the average over the individual configurations of a simulation trajectory.

The averages were computed for the last 2.5 ns of the FG alkane simulations and for the last 5 ns of the CG alkane simulations.

2. Heat of Vaporization. The heat of vaporization was calculated as the difference in intermolecular potential energy per mole between the gas phase and the liquid phase

$$\begin{aligned} \Delta H_{vap} &= V_{pot}^{intra}(g) - (V_{pot}^{intra}(l) + V_{pot}^{inter}(l)) + RT \\ &= V_{pot}^{intra}(g) - V_{pot}^{tot}(l) + RT \end{aligned} \tag{10}$$

where R is the gas constant, and T is the temperature, assuming ideal conditions for the gas phase and thus no intermolecular potential energy contribution $V_{pot}^{inter}(g)$. The intramolecular contribution in the gas phase was estimated from a stochastic dynamics (SD) simulation of one molecule using a friction coefficient $\gamma = 50\ ps^{-1}$.

3. Excess Free Energy. The excess free energy of each alkane liquid was obtained from thermodynamic integration (TI) calculations, in which the nonbonded interactions were switched off as a function of a coupling parameter λ[46] with the value of the exponent n in λ^n and $(1\text{-}\lambda)^n$ equal to 2, in the liquid phase using MD at constant volume, and in the gas phase using SD for a single molecule. No soft-core interaction was used. Twenty-one evenly distributed λ-values were used, at each of which the value of $\langle \partial H/\partial\lambda \rangle$ was obtained by averaging over 1 ns after equilibrating during 1 ns. The final configuration of the 10 ns standard MD simulation was used as initial configuration for the TI simulation at $\lambda = 0$ (full nonbonded interaction), while for the simulations at all other λ values the configuration after 200 ps of equilibration at the preceding λ-value served as initial configuration. The numerical integration of $\langle \partial H/\partial\lambda \rangle$ over λ was carried out using the trapezoid formula.

4. Free Enthalpy of Hydration. The free enthalpy of hydration of a single CG alkane in a cubic periodic box containing CG water[21] was obtained using the same TI procedure as used for the excess free energy, except that the liquid phase simulations were carried out at constant pressure and the simulation parameters of ref 21 for CG water were used.

5. Other Properties. The surface tension γ, heat capacity C_p, isothermal compressibility κ_T, thermal expansion coefficient α_T, self-diffusion constant D, and shear viscosity η of the CG n-alkanes were calculated in the same way as described in ref 47.

RESULTS AND DISCUSSION

Atomic-Level FG Simulations. Table 1 lists the experimental densities and heats of vaporization for the different alkanes as compiled in ref 33. Comparison of the values obtained from the FG MD simulations to the experimental data shows a good agreement for the density as well as for the heat of vaporization. The simulation data of ref 33 obtained for much smaller systems and shorter simulation periods could be confirmed, and the FG ensembles may serve as basic data for the CG force-field calibration, thereby complementing experimental data.

Bonded Interactions of the CG Alkane Model. The values for the bonded parameters of the CG model were estimated by matching the CG distribution to the distribution of the corresponding property obtained from the analysis of the FG-to-CG mapped FG simulation trajectories. Figure 1 shows

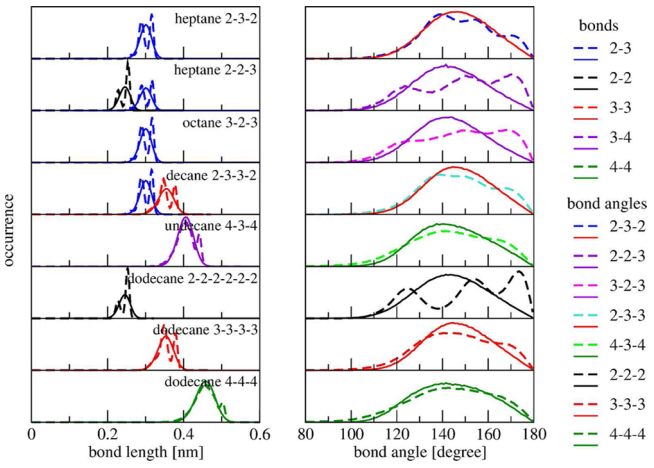


Figure 1. Bond-length and bond-angle distributions of the CG alkanes as calculated from the mapped FG MD simulation trajectories (dashed lines) and from the CG MD simulation trajectories (solid lines).

the latter together with the bond-length and bond-angle distributions as obtained from the CG alkane simulations. The distributions of the bond distances and the bond angles obtained from the mapped FG simulation trajectories show, especially for small bead sizes, multiple peaks. These occur due to the presence of different trans- or gauche-configurations in the underlying atomistic model simulation. For example, the distance of two mapped beads of bead size 2 is much smaller for the cis-configuration than for the trans-configuration of FG butane. For larger bead sizes this effect is less visible because the corresponding FG alkane fragment is longer and the

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

number of combinations of trans- and gauche-configurations with very similar bead−bead distances or angles increases, resulting in a broader and flatter distribution. It is clear that a potential energy term for a CG bond or bond-angle as described in eqs 1 and 2 cannot reproduce in detail the mapped FG distribution, but the FG and CG distributions nevertheless show large overlap. The ideal bond angles θ_0 were chosen near the maxima of the FG distribution (140−160°) which avoids linear atom configurations within the alkane chains and allows for applying a torsional-angle potential energy term, eq 3. The potential energy term of the torsional angles should not be flat, which becomes clear from Figure 2. This term is neglected in other models.[25,29,48] The probability of small torsional angles is higher than for angles around 180°. The computed torsional-angle distributions from the CG simulations are in good agreement with torsion-angle distributions from the mapped FG simulations. The values of the parameters of bonded interactions for CG alkanes are listed in Table 2.

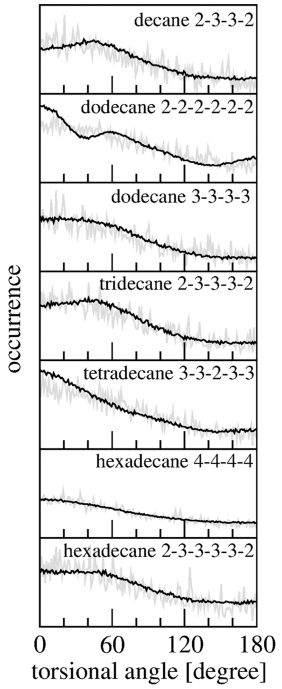


Figure 2. Torsional-angle distributions of the CG alkanes as calculated from the mapped FG MD simulation trajectories (black lines) and from the CG MD simulation trajectories (gray lines). For hexadecane in the lowest panel, only distributions of 3−3−3−3 torsional-angle type are shown.

Lennard-Jones Interactions of the CG Alkane Model. The calculation of the $C_{12}(i, j)$ and $C_6(i, j)$ parameters of the LJ potential energy from the mapped interbead LJ potential energy distribution shows the need to distinguish the CG particles depending on their position in the alkane chain, especially for larger bead sizes, see Table S1 of the Supporting Information. Therefore, we defined six types of CG beads, namely C2E, C2M, C3E, C3M, C4E, and C4M, where CX indicates the mapping of X FG carbon united atoms to one bead, and the indices E (end) and M (middle, i.e. not terminal) denote the position of the bead within an alkane chain.

A first CG alkane simulation showed poor agreement of the density and the heat of vaporization for the different alkanes (data not shown). Thus, the LJ parameters obtained from the mapped bead−bead nonbonded energy distributions needed to be refined with respect to reproduction of the experimental density and heat of vaporization. Results of such a refinement for CG *n*-alkanes are given in Tables S2−S29 in the Supporting Information which illustrate the behavior of the density and the heat of vaporization, respectively, while altering the LJ ε and σ parameters.

This refinement parametrization yielded scaling factors f_ε of sizes 1.027, 0.974, 0.824 for bead sizes 2, 3, and 4 for end beads to modify the initial ε and the factors f_σ of sizes 1.014, 1.095, 1.199 for σ (Table S30). For middle beads, the averaged scaling factors f_ε are 1.0542, 1.0660, 1.0075 with respect to bead sizes 2, 3, and 4, and factors f_σ are 0.99, 1.03, and 1.09, respectively (Table S31). These factors seem to be small but yet induce sizable changes in the C_{12} and C_6 LJ parameters due to the powers of 12 and 6 in eq 4. The final LJ parameters are listed in Table 3 and differ in some cases by more than a factor of 2 from those in Table S1 of the Supporting Information. The corresponding LJ potential energies as a function of *r* are shown in Figure 3, which also shows the potential energy of the mapped CG alkanes based on the FG simulation trajectories and the LJ potential energy for the FG CH_2 and CH_3 united atoms. The single-bead-size CG alkanes, i.e. coarse-grained by

Table 3. Coarse-Grained Particle Types and the Corresponding Lennard-Jones Interaction Parameters for Use in Eq 4

				Lennard-Jones parameters	
particle name	bead size	fine-grained equivalent	mass [u]	C_{12} [10^{-4} kJ mol^{-1} nm^{12}]	C_6 [10^{-2} kJ mol^{-1} nm^6]
Parameters of the CG Alkane-CG Alkane LJ Interaction					
C2M	2	2 CH_2	28.054	2.33569	3.27838
C2E	2	1 CH_3, 1 CH_2	29.062	3.00128	4.34369
C3M	3	3 CH_2	42.081	7.35078	8.12216
C3E	3	1 CH_3, 2 CH_2	43.089	10.88048	10.24756
C4M	4	4 CH_2	56.108	17.59416	15.56401
C4E	4	1 CH_3, 3 CH_2	57.116	32.14380	20.42999
Parameters of the CG Alkane-CG Water[21] LJ Interaction					
C2M	2	2 CH_2	28.054	3.02004	3.27838
C2E	2	1 CH_3, 1 CH_2	29.062	2.69815	4.34369
C3M	3	3 CH_2	42.081	10.8865	8.12216
C3E	3	1 CH_3, 2 CH_2	43.089	11.00017	10.24756
C4M	4	4 CH_2	56.108	25.8986	15.56401
C4E	4	1 CH_3, 3 CH_2	57.116	35.71176	20.42999

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

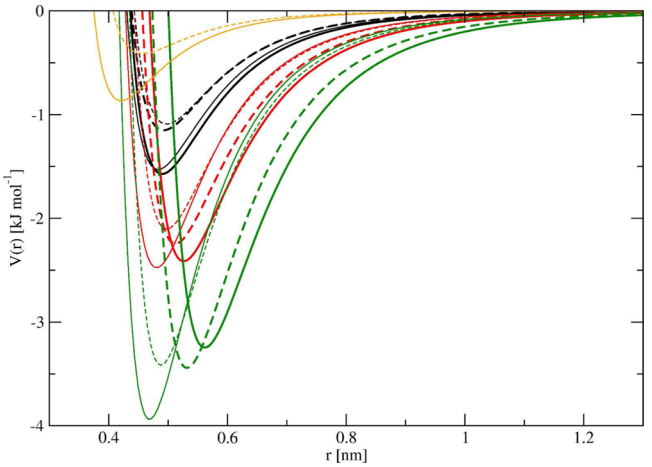


Figure 3. Lennard-Jones potential energy $V(r)$ as a function of the interparticle distance r. Thick lines: potential energy of the parametrized CG model for the CXE (X = 2, 3, 4) beads (solid line) and CXM (X = 2, 3, 4) beads (dashed line). Thin lines: the CG potential energy calculated from the mapped FG simulation trajectories. Colors indicate the bead size: 2 (black), 3 (red), and 4 (green). The Lennard-Jones potential energy of the FG CH_2 (dashed) and CH_3 (solid) particles are shown in orange.

beads all of the same bead size, show a good agreement with experimental densities and ΔH_{vap} values for all chain lengths (Table 4 and Figure 4 (a,b)).

Table 4. Density ρ and Heat of Vaporization ΔH_{vap} for Coarse-Grained Alkanes at 298 K and 1 atm, unless Indicated Otherwise[c]

		ρ [kg m^{-3}]		ΔH_{vap} [kJ mol^{-1}]	
alkane	bead sizes	exp	calc	exp	calc
butane	2-2	573.0[a]	572.2	21.62	21.64
pentane	2-3	626.2[b]	623.0	26.43	26.49
hexane	2-2-2	660.6	655.5	31.55	31.87
	3-3		660.0		31.50
heptane	2-3-2	679.5	676.2	36.55	36.87
	2-2-3		683.2		36.72
octane	2-2-2-2	698.6	701.1	41.49	41.40
	4-4		698.0		41.38
	3-2-3		703.0		41.60
nonane	3-3-3	719.2[b]	719.0	46.44	46.45
decane	2-2-2-2-2	726.6	730.0	51.37	51.76
	2-3-3-2		725.5		51.80
undecane	4-3-4	740.2[b]	736.3	56.33	56.26
dodecane	2-2-2-2-2-2	749.5[b]	749.6	61.29	61.08
	3-3-3-3		750.1		61.49
	4-4-4		747.9		61.32
tridecane	2-3-3-3-2	756.4[b]	752.7	66.23	66.67
tetradecane	2-2-2-2-2-2-2	759.6[b]	763.5	71.17	70.85
	3-3-2-3-3		762.1		70.90
pentadecane	3-3-3-3-3	768.5[b]	768.0	76.15	75.82
hexadecane	2-2-2-2-2-2-2-2	770.1	773.3	81.09	80.22
	4-4-4-4		772.5		80.80
	2-3-3-3-3-2		768.5		80.79
heptadecane	3-3-3-2-3-3	778.0[b]	774.5	86.19	84.94

[a]Experimental density measured at a pressure >1 atm. [b]Experimental density measured at a temperature of 293 K. [c]Experimental values were taken from ref 51.

The multibead-size alkanes, which were not used in the nonbonded force-field parametrization, show a good agreement with experiment for the density and heat of vaporization (Table 4), indicating that the standard geometric combination rules for the Lennard-Jones interactions (eq 5) as used in FG force fields, can be used in the CG alkane model as well.

The free enthalpies of hydration of the CG alkanes in CG water calculated using the LJ parameters in the upper part of Table 3 for CG-alkane beads, the LJ parameters of the CG-water beads,[21] and the geometric combination rule eq 5 did not match the experimental values. This was not very surprising because, also at the atomic level, the interaction of pairs of apolar atoms requires different C_{12} parameters than the interaction of pairs of apolar and polar atoms.[10,33] Therefore, a separate set of CG LJ parameters for alkane CG beads was derived for the use in combination with the LJ parameters of the CG water model in eq 5. These parameters are given in the lower part of Table 3, and they provide a good agreement of the free enthalpy of hydration for the set of alkanes with experimental values, see Table 5. Although only pure-bead-size alkanes were used in the parametrization, the calculated free enthalpy of hydration of multibead-size alkanes also agrees with experimental values quite well, sometimes even better than those of pure-bead-size alkanes (Table 5 and Figure 4 (c) in Supporting Information).

Structural Analysis of the CG Alkanes. The intermolecular radial distribution functions (RDFs) from end-bead to end-bead, $g_{ee}(r)$, end-bead to middle-bead, $g_{em}(r)$, and middle-bead to middle-bead, $g_{mm}(r)$, of CG dodecane are shown in Figure 5. They are representative for all single-bead-size CG alkanes which look similar for different alkane chain lengths. All three different RDFs calculated from the mapped FG simulation trajectories of FG dodecane show about the same level of structure. The first peaks of all three RDFs are as expected shifted to longer bead−bead distances for increasing bead size. These mapped RDFs are to be compared to the RDFs obtained from the singe-bead-size CG dodecane simulations of bead sizes 2 to 4. For bead size 2, the $g_{ee}(r)$ are similar to the corresponding mapped RDF, while $g_{em}(r)$ and $g_{mm}(r)$, both involving middle-beads, show a small bump around 0.7 nm. $g_{ee}(r)$ of bead size 3 has a too high and too narrow first peak but looks fine for longer distances, while $g_{em}(r)$ and $g_{mm}(r)$ have a small bump around 0.8 nm and a too high first peak, although not as much as the one in $g_{ee}(r)$. The RDFs from bead size 4 are too structured, as already observed in an earlier study of a CG alkane model with bead size 4.[29] The narrow and high first peaks in all CG RDFs of all bead sizes, except $g_{mm}(r)$ for bead size 2, may be a result of the deep LJ potential energy function (Figure 3), which is needed to reproduce the experimental density and heat of vaporization using the CG model, and could not be resolved as stated earlier.[29] Another possibility to make the first peaks broader and lower would be to choose another interaction function than the (12/6) LJ potential energy. However, a coarse-grained model using a Morse potential energy function shows structural behavior very similar to our model including the bump around 0.8 nm for bead size 3 and generates too much structure for bead size 4.[27]

The intramolecular end-to-end distributions of the alkanes calculated from the CG alkane simulations of nonane, dodecane, and pentane are shown in Figure 6 and compared to the corresponding mapped end-to-end distance distributions computed from the FG alkane simulation trajectories. Except

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

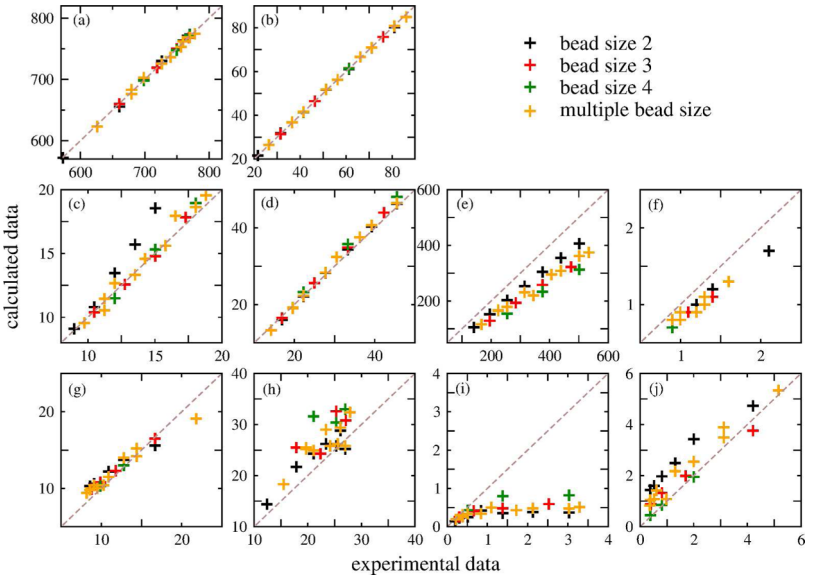


Figure 4. Comparison of experimental data and calculated data in terms of different bead size and various properties: (a) density (in kg m^{-3}); (b) heat of vaporization (in kJ mol^{-1}); (c) free enthalpy of hydration (in kJ mol^{-1}); (d) excess free energy (in kJ mol^{-1}); (e) heat capacity (in J K^{-1} mol^{-1}); (f) thermal expansion coefficient (in 10^{-3} K^{-1}); (g) isothermal compressibility (in 10^{-5} atm^{-1}); (h) surface tension (in mN m^{-1}); (i) shear viscosity (in cp); (j) self-diffusion coefficient (in 10^{-5} cm^2 s^{-1}).

Table 5. Gibbs Free Energy of Solvation ΔG_{hyd} of CG Alkanes in CG Water[21] and Excess Free Energies ΔF_{exc} Calculated from the CG Alkane Simulations with the Corresponding Experimental Values[52]

		ΔG_{hyd} (kJ mol^{-1})		ΔF_{exc} (kJ mol^{-1})	
alkane	bead sizes	exp	calc	exp	calc
butane	2-2	8.98	9.09		10.09
pentane	2-3	9.74	9.56	14.23	13.27
hexane	2-2-2	10.49	10.80	16.96	16.01
	3-3		10.39		16.48
heptane	2-3-2	11.25	10.53	19.64	19.01
	2-2-3		11.45		19.21
octane	2-2-2-2	12.01	13.47	22.29	22.14
	4-4		11.49		23.27
	3-2-3		12.66		22.31
nonane	3-3-3	12.76	12.58	25.02	25.62
decane	2-2-2-2-2	13.52	15.70	27.74	28.26
	2-3-3-2		13.33		28.32
undecane	4-3-4	14.27	14.58	30.54	32.40
dodecane	2-2-2-2-2-2	15.03	18.55	33.32	34.36
	3-3-3-3		14.79		34.87
	4-4-4		15.31		35.77
tridecane	2-3-3-3-2	15.79	15.60	36.30	37.54
tetradecane	2-2-2-2-2-2-2	16.54	21.37	39.23	40.34
	3-3-2-3-3		17.94		40.72
pentadecane	3-3-3-3-3	17.30	17.83	42.27	44.00
hexadecane	2-2-2-2-2-2-2-2	18.05	24.10	45.51	46.36
	4-4-4-4		18.95		48.15
	2-3-3-3-3-2		18.64		46.54
heptadecane	3-3-3-2-3-3	18.81	19.57		49.59

for bead size 2, the distributions show a maximum at around the same positions as the mapped distributions but with a higher probability. As for the first peaks in the RDFs, the CG models show narrower distributions than the FG mapped ones, which is not surprising considering the lack of internal flexibility of a single CG bead compared to 2 to 4 FG atoms. The short

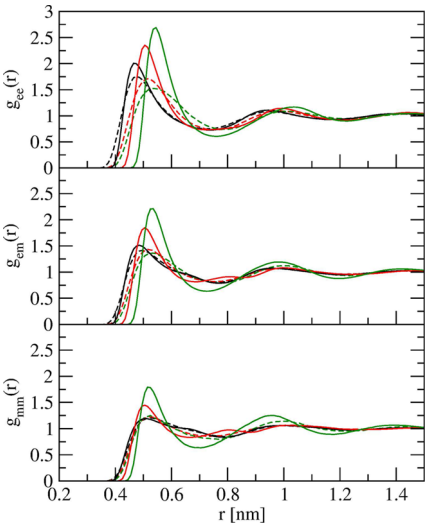


Figure 5. End−end (ee) bead, end−middle (em) bead, and middle−middle (mm) bead intermolecular radial distribution function $g(r)$ of CG dodecane. The functions are calculated from the mapped FG simulation trajectories (dashed) as well as from the CG simulation trajectories (solid). Colors indicate the bead size: 2 (black), 3 (red), and 4 (green).

end-to-end distances are better reproduced than the long-distance end-to-end distances. This may be the price paid for using an ideal bond angle smaller than 180° in the bond-angle potential energy term.

Excess Free Energy of the CG Alkanes. The excess free energies calculated from the CG alkane simulation trajectories show good agreement to values obtained from experiment (Table 5 and Figure 4 (d)), much better than for a model described and tested earlier.[29]

Other Properties. The heat capacity C_p, thermal expansion coefficient α_T, isothermal compressibility κ_T, surface tension γ, shear viscosity η, and self-diffusion constant D of the CG *n*-

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

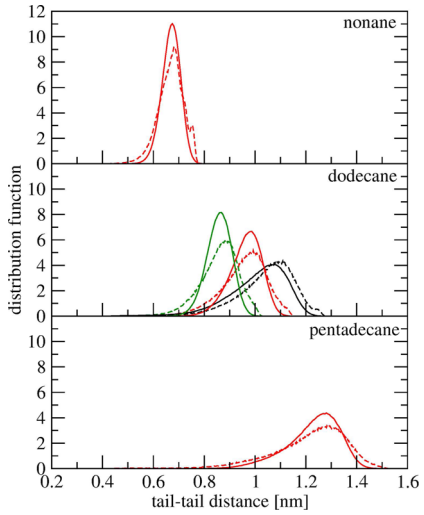


Figure 6. Intramolecular end-to-end distance distributions for nonane, dodecane, and pentadecane as calculated from the mapped FG alkane simulation trajectories (dashed) and from the CG alkane simulations (solid). Black: bead size 2; red: bead size 3; green: bead size 4.

alkanes were calculated and are shown in Tables 6 and 7 and Figure 4 (e-j).

For heat capacity, although the calculated results are always lower than the experimental data, the regular pattern of the data agrees with that of the experimental data, i.e. the heat capacity increases with increasing alkane chain length. The heat capacity is also dependent on the bead size. The larger the bead size, the smaller the heat capacity. The thermal expansion coefficient and isothermal compressibility for the CG model agree with the experimental data. The calculated surface tension is almost always a bit bigger than the experimentally measured one. The calculated shear viscosities are much lower than the experimental values and are dependent on the bead size: the larger the bead, the bigger the shear viscosity. The self-diffusion coefficients for the CG model are also clearly dependent on the bead size, as demonstrated by the example of dodecane. The larger the bead size, the smaller the diffusion constant.

SUMMARY AND CONCLUSION

A supra-atomic coarse-grained (CG) model for CG alkanes in the liquid phase has been parametrized based on experimental thermodynamic data and configurational ensembles obtained from atomic-level fine-grained (FG) simulations using the GROMOS 45A3 biomolecular force field. It consists of 6 types of coarse-grained beads or particles representing 2 to 4 fine-grained atoms, and it distinguishes between beads at the end and in the middle of an alkane chain.

Bonded interaction terms of the form of the fine-grained GROMOS 45A3 force field[33] have been used, and the force-field parameters, i.e. force constants and ideal bond lengths, bond angles, and torsional angles, were obtained from an analysis of the distributions mapped onto the CG degrees of freedom of the corresponding property as calculated from FG alkane simulation trajectories. In contrast to other CG models[25,28] our model uses an ideal bond angle θ_0 smaller than 180° and contains a torsional dihedral-angle potential energy term.

Initial Lennard-Jones nonbonded interaction parameters were obtained from an analysis of FG simulations in terms of

Table 6. Heat Capacity C_p, Thermal Expansion Coefficient α_T, and Isothermal Compressibility κ_T Calculated from the CG Alkane Simulations at 298 K and 1 atm with the Corresponding Experimental Values[51,53]

		C_p (J K^{-1} mol^{-1})		α_T (10^{-3} K^{-1})		κ_T (10^{-5} atm^{-1})	
alkane	bead sizes	exp	calc	exp	calc	exp	calc
butane	2-2	140.9	105.4	2.1[a]	1.7		27.3
pentane	2-3	167.2	116.3	1.6	1.3	21.8	19.1
hexane	2-2-2	195.6	152.6	1.4	1.2	16.7	15.6
	3-3		128.8		1.1		16.5
heptane	2-3-2	224.7	166.8	1.3	1.1	14.4	14.2
	2-2-3		165.3		1.0		15.2
octane	2-2-2-2	254.6	202.6	1.2	1.0	12.8	13.7
	4-4		154.8		0.9		13.0
	3-2-3		178.6		0.9		14.0
nonane	3-3-3	284.4	193.3	1.1	0.9	11.8	12.3
decane	2-2-2-2-2	314.4	253.5	1.0	0.9	10.9	12.2
	2-3-3-2		230.9		0.9		11.5
undecane	4-3-4	344.9	219.7	1.0	0.8	10.3	10.4
dodecane	2-2-2-2-2-2	375.8	304.4	0.9	0.8	9.9	10.6
	3-3-3-3		258.0		0.8		10.8
	4-4-4		233.4		0.8		10.3
tridecane	2-3-3-3-2	406.7	295.2	0.9	0.8	9.5	10.0
tetradecane	2-2-2-2-2-2-2	438.5	354.7	0.9	0.8	9.1	10.6
	3-3-2-3-3		308.2		0.8		10.4
pentadecane	3-3-3-3-3	473.7	322.3	0.9	0.8	8.8	9.9
hexadecane	2-2-2-2-2-2-2-2	501.6	406.6	0.9	0.8	8.6	10.3
	4-4-4-4		313.3		0.7		9.8
	2-3-3-3-3-2		361.5		0.8		9.8
heptadecane	3-3-3-2-3-3	534.7	374.4	0.9	0.8	8.2	9.4

[a]Measured at 300 K.

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 7. Surface Tension γ, Shear Viscosity η, and Self-Diffusion Coefficient D Calculated from the CG Alkane Simulations at 298 K and 1 atm with the Corresponding Experimental Values[51,54−60]

		γ (mN m^{-1})		η (cp)		D (10^{-5} cm^2 s^{-1})	
alkane	bead sizes	exp	calc	exp	calc	exp	calc
butane	2-2	12.4[a]	14.4	0.20[b]	0.14		7.93
pentane	2-3	15.5	18.3	0.22	0.19	5.16[a]	5.34
hexane	2-2-2	17.9	21.7	0.30	0.23	4.21	4.73
	3-3		25.5		0.25		3.76
heptane	2-3-2	19.7	25.2	0.39	0.26	3.12	3.89
	2-2-3		25.5		0.31		3.49
octane	2-2-2-2	21.1	24.3	0.51	0.25	2.00	3.43
	4-4		31.6		0.43		1.95
	3-2-3		24.9		0.33		2.55
nonane	3-3-3	22.4	24.3	0.66	0.41	1.70	1.99
decane	2-2-2-2-2	23.4	26.2	0.84	0.42	1.31	2.49
	2-3-3-2		29.0		0.34		2.17
undecane	4-3-4	24.2	25.8	1.10	0.50	0.98	1.08
dodecane	2-2-2-2-2-2	25.3[a]	25.8	1.38	0.35	0.81	1.97
	3-3-3-3		32.6		0.48		1.32
	4-4-4		30.4		0.80		0.85
tridecane	2-3-3-3-2	25.6	26.2	1.72	0.43	0.61	1.42
tetradecane	2-2-2-2-2-2-2	26.1	28.8	2.13	0.38	0.52	1.60
	3-3-2-3-3		29.4		0.48		1.08
pentadecane	3-3-3-3-3	27.1[a]	30.8	2.53	0.59	0.40	0.88
hexadecane	2-2-2-2-2-2-2-2	27.0	25.2	3.03	0.37	0.38	1.44
	4-4-4-4		33.0		0.82		0.45
	2-3-3-3-3-2		25.8		0.47		1.04
heptadecane	3-3-3-2-3-3	27.9[a]	32.4	3.29[c]	0.51	0.36	0.83

[a]Measured at 293 K. [b]Measured at 273 K. [c]Measured at 303 K.

mapped CG bead−bead distances and energies. Since these parameters could not reproduce experimental densities and heats of vaporization, they were refined with respect to reproduction of these quantities. The refined model shows good agreement with experimental data for the density and heat of vaporization, independent of the chain length of the alkane.

Use of large bead sizes in a CG model would result in a larger speed-up of the simulations compared to FG models. However, the liquid becomes inevitably more structured with increasing bead size as observed from the bead−bead radial distribution functions calculated from the CG alkane simulations. Beads of size 2 and 4 can be used in combination with beads of size 3 allowing for CG alkane simulations of arbitrary chain length. Although not used in the parametrization of the CG model, these multibead-size CG alkanes show good agreement between simulation and experiment with respect to the density, heat of vaporization, excess free energy, and free enthalpy of hydration (Tables 4 and 5, Figure 4).

A separate set of CG Lennard-Jones parameters for CG alkanes to be used in combination with the CG Lennard-Jones parameters of the polarizable CG water model of ref 21 and a geometric combination rule was needed to obtain free enthalpies of hydration matching experimental values. The use of a polarizable model in conjunction with a nonpolarizable one for different molecules in the condensed phase has been investigated for water[49] and other molecules. It yields good properties for such mixtures.[50] The proposed CG alkane model could be made polarizable by adding charge-on-spring sites to the CG beads of the alkanes. The CG Lennard-Jones parameters for alkane-water interactions are then to be recalibrated, as done in ref 47.

The proposed CG model was further tested by calculating the excess free energies, self-diffusion coefficients, surface tension, heat capacity, isothermal compressibility, thermal expansion coefficient, self-diffusion constant, and shear viscosity of alkanes. The experimental values for these quantities were rather well reproduced.

The proposed CG model offers a solid basis from where a CG model for lipids can be developed. In that case, the use of bead size 3 seems appropriate in order to avoid too thick spherical lipid tails, while using bead size 2 to get to the proper number of carbon atoms represented.

■ ASSOCIATED CONTENT

Supporting Information

Coarse-grained bead or particle types and the corresponding Lennard-Jones interaction parameters as calculated from the mapping of the FG alkane simulation trajectories onto CG degrees of freedom are given in Table S1. Refinement of Lennard-Jones parameters for different bead types with respect to reproduction of the experimental density and heat of vaporization is shown in Tables S2−S29. Scaling factors of the Lennard-Jones parameters obtained from CG simulations of liquid alkanes for the CXE beads are given in Table S30 and for the CXM beads in Table S31. Scaling factors of the Lennard-Jones C_{12} parameter obtained from CG simulations of individual alkanes in CG water for the beads of different sizes are given in Tables S32 and S33. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00295.

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

AUTHOR INFORMATION

Corresponding Author

*E-mail: wfvgn@igc.phys.chem.ethz.ch.

Author Contributions

†These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was financially supported by grant number 200020-137827 of the Swiss National Science Foundation and by grant number 228076 of the European Research Council (ERC), which is gratefully acknowledged. The authors thank Dr. Victor Holanda Rusu for useful discussions.

REFERENCES

(1) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267*, 585−590.

(2) Karplus, M. Molecular dynamics of biological macromolecules: A brief history and perspective. *Biopolymers* **2003**, *68*, 350−358.

(3) van Gunsteren, W. F.; Dolenc, J. Thirty-five years of biomolecular simulation: development of methodology, force fields and software. *Mol. Simul.* **2012**, *38*, 1271−1281.

(4) Hünenberger, P. H.; van Gunsteren, W. F. Empirical classical force fields for molecular systems. In *Potential Energy Surfaces*; Sax, A. F., Ed.; Springer: Berlin, Heidelberg, 1999; pp 177−214.

(5) Weiner, P. K.; Kollman, P. A. Amber - assisted model-building with energy refinement - a general program for modeling molecules and their interactions. *J. Comput. Chem.* **1981**, *2*, 287−303.

(6) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(7) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(8) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(9) van Gunsteren, W. F.; Karplus, M. Effect of constraints on the dynamics of macromolecules. *Macromolecules* **1982**, *15*, 1528−1544.

(10) 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. Biophys. Lett.* **2011**, *40*, 843−856.

(11) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Biomolecular modeling: goals, problems, perspectives. *Angew. Chem., Int. Ed.* **2006**, *45*, 4064−4092.

(12) Sherwood, P.; Brooks, B. R.; Sansom, M. S. P. Multiscale methods for macromolecular simulations. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630−640.

(13) 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.

(14) Cascella, M.; Dal Peraro, M. Challenges and perspectives in biomolecular simulations: from the atomistic picture to multiscale modeling. *Chimia* **2009**, *63*, 14−18.

(15) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems. *Annu. Rev. Phys. Chem.* **2011**, *62*, 41−64.

(16) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423−12430.

(17) van Gunsteren, W. F.; Berendsen, H. J. C. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* **1977**, *34*, 1311−1327.

(18) Levitt, M.; Lifson, S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.* **1969**, *46*, 269−279.

(19) Müller, M.; Katsov, K.; Schick, M. Biological and synthetic membranes: what can be learned from a coarse-grained description? *Phys. Rep.* **2006**, *434*, 113−176.

(20) Huang, W.; van Gunsteren, W. F. Challenge of representing entropy at different levels of resolution in molecular simulation. *J. Phys. Chem. B* **2015**, *119*, 753−763.

(21) Riniker, S.; van Gunsteren, W. F. A simple, efficient polarizable coarse-grained water model for molecular dynamics simulations. *J. Chem. Phys.* **2011**, *134*, 084110.

(22) Allison, J. R.; Riniker, S.; van Gunsteren, W. F. Coarse-grained models for the solvents dimethyl sulfoxide, chloroform, and methanol. *J. Chem. Phys.* **2012**, *136*, 054505.

(23) Riniker, S.; Eichenberger, A. P.; van Gunsteren, W. F. Structural effects of an atomic-level layer of water molecules around proteins solvated in supra-molecular coarse-grained water. *J. Phys. Chem. B* **2012**, *116*, 8873−8879.

(24) Riniker, S.; Eichenberger, A. P.; van Gunsteren, W. F. Solvating atomic level fine-grained proteins in supra-molecular level coarse-grained water for molecular dynamics simulations. *Eur. Biophys. J. Biophys. Lett.* **2012**, *41*, 647−661.

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

(26) Orsi, M.; Haubertin, D. Y.; Sanderson, W. E.; Essex, J. W. A quantitative coarse-grain model for lipid bilayers. *J. Phys. Chem. B* **2008**, *112*, 802−815.

(27) Chiu, S. W.; Scott, H. L.; Jakobsson, E. A coarse-grained model based on morse potential for water and *n*-alkanes. *Biophys. J.* **2010**, *98*, 573a−574a.

(28) 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.

(29) Baron, R.; Trzesniak, D.; de Vries, A. H.; Elsener, A.; Marrink, S. J.; van Gunsteren, W. F. Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models. *ChemPhysChem* **2007**, *8*, 452−461.

(30) Huang, W.; Riniker, S.; van Gunsteren, W. F. Rapid sampling of folding equilibria of beta-peptides in methanol using a supramolecular solvent model. *J. Chem. Theory Comput.* **2014**, *10*, 2213−2223.

(31) Poger, D.; Mark, A. E. On the validation of molecular dynamics simulations of saturated and *cis*-monounsaturated phosphatidylcholine lipid bilayers: a comparison with experiment. *J. Chem. Theory Comput.* **2010**, *6*, 325−336.

(32) Schuler, L. D.; van Gunsteren, W. F. On the choice of dihedral angle potential energy functions for *n*-alkanes. *Mol. Simul.* **2000**, *25*, 301−319.

(33) Schuler, L. D.; Daura, X.; Van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **2001**, *22*, 1205−1218.

(34) Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; van Gunsteren, W. F. Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput. Phys. Commun.* **2012**, *183*, 890−903.

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(35) Kunz, A. P.; Allison, J. R.; Geerke, D. P.; Horta, B. A.; Hünenberger, P. H.; Riniker, S.; Schmid, N.; van Gunsteren, W. F. New functionalities in the GROMOS biomolecular simulation software. *J. Comput. Chem.* **2012**, *33*, 340−353.
(36) Eichenberger, A. P.; Allison, J. R.; Dolenc, J.; Geerke, D. P.; Horta, B. A. C.; Meier, K.; Oostenbrink, C.; Schmid, N.; Steiner, D.; Wang, D. Q.; van Gunsteren, W. F. GROMOS++ software for the analysis of biomolecular simulation trajectories. *J. Chem. Theory Comput.* **2011**, *7*, 3379−3390.
(37) van Gunsteren, W. F. GROMOS. The GROMOS software package and force fields can be downloaded from this Web site. http://www.gromos.net (accessed January 23, 2015).
(38) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(39) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(40) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(41) Hockney, R. W.; Eastwood, J. W. *Computer simulation using particles*; McGraw-Hill: New York, 1981.
(42) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A generalized reaction field method for molecular-dynamics simulations. *J. Chem. Phys.* **1995**, *102*, 5451−5459.
(43) Heinz, T. N.; van Gunsteren, W. F.; Hünenberger, P. H. Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. *J. Chem. Phys.* **2001**, *115*, 1125−1136.
(44) Harvey, S. C.; Tan, R. K. Z.; Cheatham, T. E. The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.* **1998**, *19*, 726−740.
(45) Huang, W.; Hansen, N.; van Gunsteren, W. F. On the use of a supramolecular coarse-grained model for the solvent in simulations of the folding equilibrium of an octa-beta-peptide in MeOH and H_2O. *Helv. Chim. Acta* **2014**, *97*, 1591−1605.
(46) Riniker, S.; Christ, C. D.; Hansen, H. S.; Hünenberger, P. H.; Oostenbrink, C.; Steiner, D.; van Gunsteren, W. F. Calculation of relative free energies for ligand-protein binding, solvation, and conformational transitions using the GROMOS software. *J. Phys. Chem. B* **2011**, *115*, 13570−13577.
(47) Szklarczyk, O. M.; Bachmann, S. J.; van Gunsteren, W. F. A polarizable empirical force field for molecular dynamics simulation of liquid hydrocarbons. *J. Comput. Chem.* **2014**, *35*, 789−801.
(48) Baron, R.; de Vries, A. H.; Hünenberger, P. H.; van Gunsteren, W. F. Comparison of atomic-level and coarse-grained models for liquid hydrocarbons from molecular dynamics configurational entropy estimates. *J. Phys. Chem. B* **2006**, *110*, 8464−8473.
(49) Bachmann, S. J.; van Gunsteren, W. F. On the compatibility of polarisable and non-polarisable models for liquid water. *Mol. Phys.* **2014**, *112*, 2761−2780.
(50) Bachmann, S. J.; van Gunsteren, W. F. Polarizable model for DMSO and DMSO-water mixtures. *J. Phys. Chem. B* **2014**, *118*, 10175−10186.
(51) *CRC Handbook of Chemistry and Physics*, 94th ed.; Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, 2013.
(52) Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **1984**, *81*, 2016−2027.
(53) Miyamoto, H.; Uematsu, M. (p, ρ, T) Properties for *n*-butane in the temperature range from 280 to 380 K at pressures up to 200 MPa. *J. Chem. Thermodyn.* **2007**, *39*, 588−593.
(54) Wu, J. N.; Shan, Z. J.; Asfour, A. F. A. Viscometric properties of multicomponent liquid n-alkane systems. *Fluid Phase Equilib.* **1998**, *143*, 263−274.
(55) Kumagai, A.; Tomida, D.; Yokoyama, C. Measurements of the liquid viscosities of mixtures of *n*-butane, *n*-hexane, and *n*-octane with squalane to 30 MPa. *Int. J. Thermophys.* **2006**, *27*, 376−393.
(56) Hopfe, D. *Data Compilation of FIZ. CHEMIE*; Germany, 1990; p 27.
(57) Liessman, G.; Schmidt, W.; Reiffarth, S. *Data Compilation of the Saechsische Olefinwerke Boehlen*; Germany, 1995; p 61.
(58) Liessman, G.; Schmidt, W.; Reiffarth, S. *Data Compilation of the Saechsische Olefinwerke Boehlen*; Germany, 1995; p 73.
(59) Liessman, G.; Schmidt, W.; Reiffarth, S. *Data Compilation of the Saechsische Olefinwerke Boehlen*; Germany, 1995; p 86.
(60) Douglass, D. C.; McCall, D. W. Diffusion in paraffin hydrocarbons. *J. Phys. Chem.* **1958**, *62*, 1102−1107.

DOI: 10.1021/acs.jctc.5b00295
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX