Conditional reversible work coarse-grained models of molecular liquids with Coulomb electrostatics – a proof of concept study on weakly polar organic molecules

Gregor Deichmann* and Nico F. A. van der Vegt

Technische Universität Darmstadt, Eduard-Zintl-Institut für Anorganische und Physikalische Chemie, Alarich-Weiss-Straße 10, 64287 Darmstadt, Germany

E-mail: deichmann@cpc.tu-darmstadt.de

DOI: https://doi.org/10.1021/acs.jctc.7b00611

Abstract

Scale bridging simulations of soft matter rely on the availability of transferable coarse-grained models. In systematic coarse-graining approaches, core principles of statistical mechanics are used to relate the coarse-grained models to the underlying molecular interactions. The conditional reversible work (CRW) method provides effective, nonbonded pair potentials by means of computing coupling free energies between mapped chemical groups. This method has so far been used almost exclusively for systems composed of apolar organic molecules, but additional challenges arise when developing coarse-grained models for polar molecules in which (long-ranged) electrostatic interactions are important. Herein, we present a modified formulation of the CRW method where we divide the effective interaction potential into van-der-Waals

and electrostatic components. The shape of the effective electrostatic interaction justifies modeling the electrostatics using a Coulomb potential with point charges on each site that are equal to the net charge of the underlying group of atoms. We perform CRW calculations using two polar molecules as test cases (an ether (1,2-dimethoxyethane) and an ester (ethylpropionate)). The results of subsequent liquid state simulations indicate that the coarse-grained models obtained by the new method are of similar quality with respect to representability and thermodynamic transferability as formerly developed models for apolar systems.

Introduction

Soft matter research often requires the study of systems and processes on time and length scales that are too large to be modeled by atomistic models. The use of coarse-grained (CG) models enables us to overcome some of these limitations by reducing the number of degrees of freedom in the system and thereby softening the potential energy surface which leads to inherently faster dynamics.[1] A variety of methods have been published in the literature that follow the systematic coarse-graining approach.[2–11] In systematic coarse-graining methods, interactions between sites in the coarse-grained model are based on an underlying fine-grained model in a bottom-up manner. The resulting coarse-grained models have been used for the study of a variety of soft matter systems.[12–16] These methods can be subdivided into two classes: structure based (IBI,[6] IMC[7]) and force-matching (MS-CG,[5] EF-CG,[4] CRW[3]) methods. In structure based coarse-graining, one tries to parametrize the model in such a way that the liquid structure and additional properties are reproduced by the CG model. Force-based methods on the other hand generate a CG potential based directly on the interactions in the fine-grained model. The conditional reversible work (CRW) method developed by Brini et al. is of the latter kind,[3] and it is unique because it can be applied by sampling only two molecules in a vacuum environment, thereby providing a CG model at minimal computational cost.

Other CG models have been parametrized in a rather top-down approach with the aim of reproducing e.g. vapor-liquid coexistence curves,[17] polymer glass transition properties[18] or bulk partition free energies.[19] The use of CG models is in many cases restricted to state points close to the state point at which they were parametrized. Transferability, i.e. the capability of a model of correctly representing state points beyond the immediate vicinity of the state point of its parametrization, is a desired feature of CG models and several approaches have been published to achieve transferability. These include the use of temperature-dependent potentials,[17,18,20–22] multistate sampling techniques[23,24] and other approaches.[8,25,26]

We have shown in the past that the CRW coarse-graining method provides CG models that are transferable in a rather broad range of temperatures and that these models represent liquid phase structural and thermophysical properties as well as liquid-vapor and liquid-solid interfacial properties in good agreement with experiments and fine-grained simulations.[3,14,27] The CRW method has been used to develop CG models for alkanes, polymers and graphene surfaces.[3,12,14,27–29] It has so far not been applied for polar or ionic systems, and in its original formulation it is not well suited for the application to such systems. Herein we present a modified formulation of the CRW method that allows for a separate treatment of electrostatic and Van-der-Waals (vdW) interactions in the CG model. This modified CRW method is subsequently tested on two model systems (an ether and an ester). We would like to point out that it is not the focus of this work to provide CG models whose properties match experimental data. Rather, we would like to show in a "proof-of-principle" manner that the method described herein allows to obtain transferable CG models from underlying fine-grained reference representations of polar molecules.

Methods

The aim of the systematic coarse-graining procedure is the development of a coarse-grained (CG) model that is based on a fine-grained (FG) reference model (usually an all-atom or

united-atom model). The CG model consists of interaction sites that are related to the sites of the FG model by a mapping scheme. In this work, the center of mass mapping is used (see Computational Details).

The main challenge of the coarse-graining process is the determination of suitable non-bonded interaction potentials. In CRW models, interactions between pairs of CG sites are modeled with a free-energy-based pair potential, which corresponds to the reversible work of coupling the microscopic interactions between the atoms contained within the CG interaction sites, given a constant distance between CG sites. This free energy difference can be calculated by the use of a thermodynamic cycle (see fig. 1) from the reversible work (RW) associated with the process of moving the two interaction sites from infinity to a finite distance r with direct interactions turned on and off respectively:

$$G_{eff}(r) = W_{on}(r) - W_{off}(r) \tag{1}$$

A feature that distinguishes the CRW method from other coarse-graining methods is the fact that the CG potentials can be obtained from atomistic sampling of two molecules in vacuo. Other surroundings can in principle also be chosen,[14] but the sampling in vacuo provides a computationally efficient way to calculate the CRW-CG interaction potentials. All CRW potentials used in this work were obtained using vacuum sampling. For a more detailed description of the CRW method we refer the reader to the existing literature.[3]

When modeling apolar molecules, the resulting CG potentials will resemble a Lennard-Jones potential with its characteristic features: a steep increase of energy towards zero distance, a potential well, and an attractive tail converging to zero at short distances. The last feature of the potentials is critical for the success of the method as using the aforementioned thermodynamic cycle necessitates that $G_{eff} \approx 0$ at the largest sampled inter-site distance. For models of polar molecules the interaction potential may contain significant contributions of electrostatic interactions, especially if the mapping scheme is chosen such that interaction

sites exhibit net charge $\neq 0$. These electrostatic contributions are typically long-ranged and thus may not have converged to zero at the distance chosen as interaction cutoff. The resulting CRW potential is shifted by an unknown amount, and it is unclear whether it can serve as a physically meaningful interaction potential. Extending the sampling by including more and more distances will, of course, eventually lead to convergence but the larger cutoff length required in the CG simulation will decrease the overall performance. This can be overcome by including a switch function that will force the interactions to zero at a shorter distance value. The choice of this shorter interaction cutoff of this switch function, however, is an additional parametrization choice and may consequently require tuning of the non-bonded interaction potentials.

In this work, we propose a different approach to tackle this challenge: we separate the interaction potential into an effective Van-der-Waals (vdW) and an effective electrostatic (ES) part. This natural decomposition is also used in atomistic force fields and in the effective-force coarse-graining method.[4] With the available algorithms for the treatment of long-range electrostatics, such as particle mesh Ewald and reaction field algorithms, a short cutoff can also be used for the ES part of the interaction potential.

The decomposition of G_{eff} is performed as described in the following. We assert here that non-bonded atomistic interactions in the fine-grained force field underlying our CG model are modeled using a Van-der-Waals and an electrostatic term. We will refer to these terms as the "interaction components" of the fine-grained model. The process of "switching on" these components will be performed in two steps, as opposed to the single step procedure employed in the original CRW method. In the first step, one of the FG interaction components is turned on leading to an intermediate state. The free energy difference associated with this process will be denoted as A_Ψ. Next, the second interaction component is introduced. The free energy difference associated with this change of the FG interactions is denoted B_Φ. In our notation Ψ and Φ serve as placeholders for the interaction components that are being introduced into the model (i.e. ES and vdW). The CG interaction energy is the sum of the

two parts:

$$G_{eff} = A_\Psi + B_\Phi \tag{2}$$

The order of introduction can be chosen arbitrarily and the resulting G_{eff} will be independent of this choice. However, one should be aware of the fact that this is not true for its components A and B:

$$A_\Psi \neq B_\Psi \tag{3}$$

because the second introduction is performed under the condition that the first interaction component is already present in the system.

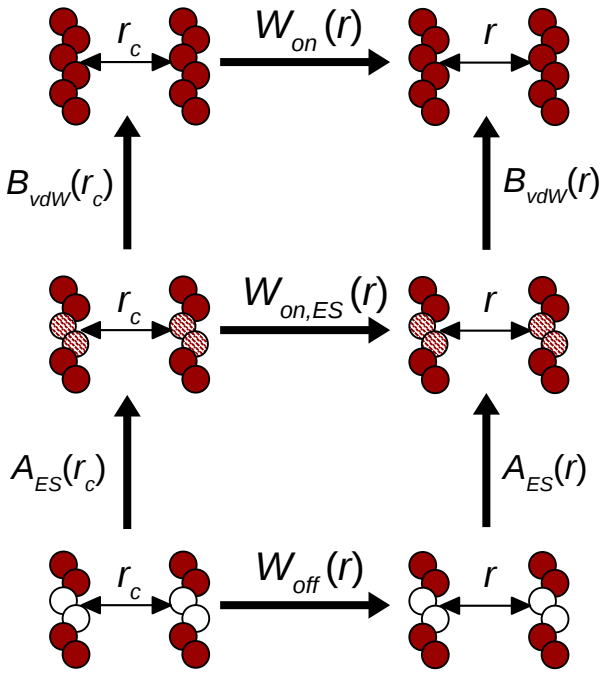


Figure 1: Thermodynamic cycles used for the calculation of $A(r)$ and $B(r)$. The two central atoms of a hexamer are merged into one site in the CG model (2:1 mapping). In the lowest part, all interactions between the two central atoms are absent; in the middle part, only the electrostatic component of the interactions is present; and in the upper part, all interactions are included. $A(r)$ and $B(r)$ are then calculated from the potentials of mean force $W(r)$ under the assumption that all $W(r_c) = 0$.

The calculation of the free energy differences is almost identical to the original single-step procedure. In addition to W_{on} and W_{off}, we calculate a RW function for the intermediate state in which interaction component Ψ is present in the system ($W_{on,\Psi}$).

The reversible work is calculated by integrating over the projection of the inter-molecular force $\mathbf{F}_{mol}$ on the unit vector $\mathbf{e}$ parallel to the distance vector connecting the mass centers of the CG sites of interest in constrained molecular dynamics (MD) simulations:

$$W(r) = -\int_{\infty}^{r} \langle \mathbf{F}_{mol} \cdot \mathbf{e} \rangle (r')\, dr' \tag{4}$$

To ensure that the distance between the CG interaction sites of interest remains constant throughout the simulation the positions of these sites (α) are constrained to their initial value by modifying the force on all atoms (i) contained within the site in the following way:

$$\mathbf{F}_{i,\alpha} = \mathbf{F'}_{i,\alpha} - \frac{m_i}{m_\alpha} \sum_{j=0}^{N_\alpha} \mathbf{F'}_{j,\alpha} \tag{5}$$

where $\mathbf{F'}_{i,\alpha}$ is the (unconstrained) force on atom i, m_i and m_α are the masses of atom i and site α respectively, and N_α is the number of atoms contained in site α. The equation of motion of atoms that are not part of the sites of interest remains unchanged.

These constrained dynamics runs are performed using four different types of interactions between the atoms within the sites of interest and the forces on the molecules are integrated to yield the corresponding reversible work:

1. All interactions present (yielding $W_{on}(r)$)
2. No interactions present (yielding $W_{off}(r)$)
3. Only electrostatics present (yielding $W_{on,ES}(r)$)
4. Only Van-der-Waals interactions present (yielding $W_{on,vdW}(r)$)

From these the CRW interaction potentials are calculated:

$$A_\Psi = W_{on,\Psi} - W_{off} \tag{6}$$

$$B_\Psi = W_{on} - W_{on,\Psi} \tag{7}$$

For the sake of clarity, we re-emphasize that only those interactions between atoms that are part of the CG sites of interest are modified; all other FG interactions remain unchanged.

This approach provides either $\{A_{ES}, B_{vdW}\}$ or $\{A_{vdW}, B_{ES}\}$ that can in principle both serve as interaction potentials in our CG simulations. It would be highly advantageous if the electrostatic component can be approximated with a Coulomb potential. Then, established long-range methods for the electrostatics can be used, avoiding a large cutoff radius in the CG simulation. We shall show below that this is indeed possible if we choose A_{ES} as a representation of the electrostatic interactions.

Computational details

The two-step CRW method is applied to two molecules, 1,2-dimethoxyethane (DXE) and ethylpropionate (EPP), mapped on a CG representation with three sites as test cases.

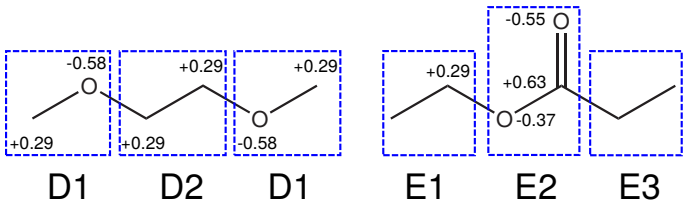


Figure 2: Mapping schemes used for the CG models. Partial charges on the united atoms used in the FG simulation are shown to illustrate that there are non-zero net charges on sites D1, D2, E1, and E2.

Figure 2 shows the mapping schemes employed for these molecules. The GROMOS $53A6_{\mathrm{OXY}}$ force field developed by Horta et al. is used as a fine-grained reference model in the CRW calculations.[30]

Calculation of non-bonded CG interaction potentials

In order to obtain the interaction potentials for the CG model, calculations are performed using the modified two-step CRW method described above.

Constrained MD runs are performed for all pairs of CG interaction sites using two molecules in vacuo. These simulations were performed using LAMMPS.[31] A time step of 1 fs is used and the total simulation time is 2 ns for each MD run. The temperature is kept at a value of 300 K using the Langevin thermostat[32] with a time constant of 0.5 ps. Distances from 0.3 nm up to a cutoff length of 1.3 nm are sampled with a 0.02 nm distance increment. The electrostatic interactions are modeled by a Coulomb potential and are screened with the experimental relative permittivity as reported in reference 30 (DXE: 7.3, EPP: 5.6) and an infinite cutoff length without long-range corrections. This leads to a closer approximation of the behavior within a liquid phase than the use of unscreened electrostatic interactions would.

It will be demonstrated below that A_{ES} resembles a Coulomb potential. Therefore, rather than A_{ES} itself, an unscreened Coulomb potential (with $\varepsilon_r = 1$) is instead applied in the CG simulations using the net charges of the interaction sites determined by the mapping (fig. 2). Long-range electrostatics are modeled using the particle-mesh Ewald (PME) algorithm.[33]

The vdW component in the CG model is modeled by B_{vdW}. In our simulations we do not use a tabulated version of B_{vdW} but instead a Morse potential, described below, fitted to the data. This choice facilitates applications of the reported CG model in which the functional form is used directly, or can be compiled in potential energy tables. A modified Morse potential with 4 parameters has been used to fit B_{vdW}:

$$U(r) = \varepsilon((1 - e^{-k_i(r-r_0)})^2 - 1) \begin{cases} i = 1 & r \leq r_0 \\ i = 2 & r > r_0 \end{cases} \tag{8}$$

For very small distances this functional form differs increasingly from the calculated values of B_{vdW}. We therefore limit our fit to distance values for which $B_{vdW} < 10$ kJ mol^{-1} to obtain a better approximation of the potential energy well and the attractive tail. The error introduced by this procedure is negligible because in an unconstrained simulation the values with considerable deviation from the fitted function are rarely sampled, and the

potential will certainly keep its repulsive character when using the functional form. In fig. 3, we motivate the choice of this functional form by comparing to a Lennard-Jones function. The Morse potential yields a very good approximation of both the excluded volume and the attractive potential well. The Lennard-Jones functional form does not yield the same degree of accuracy for both of the potentials shown in fig. 3. While the attractive region of the potential energy curve can be modeled with the Lennard-Jones functional form, the resulting core repulsion is too steep when compared to the calculated interaction energy. This result is in agreement with findings published by others.[29,34]

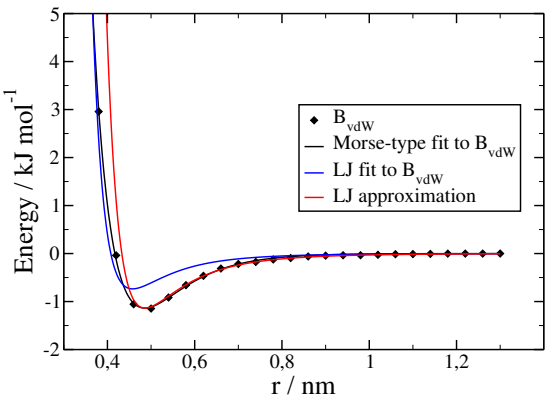


Figure 3: Different analytical approximations to B_{vdW} are shown. The modified Morse potential (black) provides the best approximation of the calculated potential. A Lennard-Jones potential obtained by the same fitting procedure (blue curve) does not match the attractive well of the potential. A Lennard-Jones potential that approximates the Morse potential (using $\sigma_{\mathrm{LJ}} = r_0/2^{\frac{1}{6}}$ and $\varepsilon_{\mathrm{LJ}} = \varepsilon$; red curve) reproduces the attractive region but does not match the excluded volume as well as the Morse-type fit.

Bonded CG interaction potentials

For DXE the intramolecular interaction potentials are calculated from Boltzmann-inversion of probability distributions gathered from a simulation run of length 100 ns of one molecule in vacuum with all nonbonded interactions excluded. The Langevin thermostat is used with a reference temperature of 300 K and a time constant of 0.5 ps. From the mapped

configurations gathered from these simulations we can calculate the potentials for the D1-D1 bond length b and the D1-D2-D1 angle θ:

$$U_B(b) = -k_B T \log \frac{\rho(b)}{b^2} + C_B \tag{9}$$

$$U_A(\theta) = -k_B T \log \frac{\rho(\theta)}{\sin(\theta)} + C_A \tag{10}$$

where ρ denotes the probability density of the respective intra-molecular degree of freedom and C is a constant chosen such that $\min(U) = 0$. The resulting potentials are shown in figure 4.

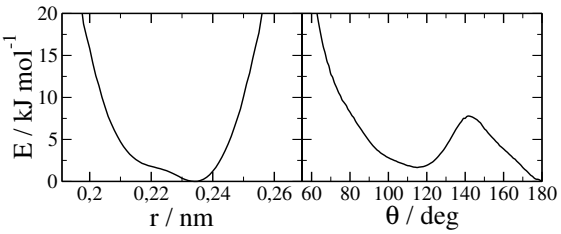


Figure 4: Intramolecular potentials used for the DXE model. Left panel: U_B. Right panel: U_A.

For EPP the same approach was attempted but as a result of the very rigid geometry of the molecule, the intramolecular potentials obtained from this calculation were too rigid to allow for a stable MD simulation at a timestep of 1 fs. We therefore constrain the bond lengths and angles of EPP to their average value using the LINCS-algorithm.[35] The average bond distances are 0.268 nm and 0.239 nm for the E1-E2 and E2-E3 bonds respectively. The average E1-E2-E3 angle is 124°.

Liquid Phase simulations

Using the FG and CG models, simulations of liquid systems are performed in a cubic simulation box with periodic boundary conditions containing 400 molecules. All liquid phase

simulations were performed with a timestep of 1 fs using the GROMACS package version 4.6.7.[36] The simulation boxes are equilibrated for 5 ns using weak-coupling algorithms for temperature and pressure.[37] After equilibration, production runs of 5 ns are performed using the Nosé-Hoover thermostat[38,39] (time constant 0.2 ps) and Parrinello-Rahman barostat[40,41] (time constant 2.5 ps; reference pressure 1 bar; compressibility $4.5 \cdot 10^{-5}$ bar^{-1}). The temperature is varied from 260 K to 340 K in increments of 20 K. In order to calculate the surface tension NVT simulations are carried out with a half empty box starting from an equilibrated NPT configuration. The same overall settings are used as for the NPT simulations.

In the FG simulations, we use the twin-range cutoff scheme of the GROMOS force field[42] using a short-range cutoff of 0.8 nm and a long-range cutoff of 1.4 nm. Neighbor lists are updated every 5 steps. A short range cutoff of 1.4 nm is used for the electrostatic interactions. Long-range electrostatic interactions are modeled using the PME algorithm implemented in GROMACS with a grid spacing of 0.12 nm and a relative tolerance of 10^{-5}.[33] No long-range corrections are applied for the Lennard-Jones dispersion contribution to the energy and pressure.

In the CG simulations, a simple cutoff of 1.3 nm is used. The Van-der-Waals interaction is simulated using a potential table generated from equation 8 with a spacing of 1 pm and linear interpolation. The short-range cutoff length for the electrostatic interactions is 1.3 nm in the CG model. The long-range part of the ES interactions is treated in the same way as in the FG simulations.

Results and Discussion

Non-bonded interaction potentials

In figure 5, we show a detailed overview of the components of the CRW interaction free energy for the pair D1-D2. This example illustrates the fact that the order of introducing the interaction components is important and why we have chosen the set $\{A_{ES}, B_{vdW}\}$ as the

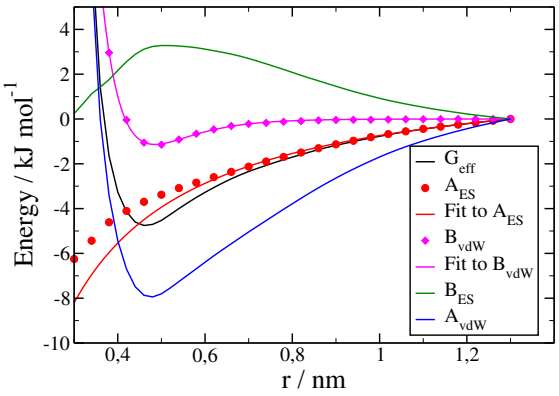


Figure 5: A detailed illustration of the interaction energies obtained using both possible introduction orders for the pair of CG sites D1 and D2. A_{ES} can be approximated by a Coulomb potential and B_{vdW} is fitted using a Morse-type potential.

final CG force field. Here B_{vdW} is fitted using eq. 8 and we can observe a good reproduction of the data by the functional form. The electrostatic component A_{ES} is compared to the Coulomb potential for point charges q_α and q_β shifted to zero at the cutoff distance:

$$U_C(r) = \frac{q_\alpha q_\beta}{4\pi\varepsilon_0\varepsilon_r}\left(\frac{1}{r} - \frac{1}{r_c}\right) \tag{11}$$

The values of q_α and q_β are set equal to the net charge on the sites here and in all following plots and are defined as:

$$q_\alpha = \sum_{j=0}^{N_\alpha} q_{j,\alpha} \tag{12}$$

where $q_{j,\alpha}$ is the partial charge on atom j in CG site α and N_α is the number of atoms contained in CG site α.

Only a good match of the calculated CRW potential to this functional form justifies the approximation to model A_{ES} with a simple Coulomb potential. The smaller the inter-site distance, the more deviation from this behavior can be expected. Indeed, fig. 5 shows that an increasing difference between the theoretical form and the CRW potential is observed. However, the overall agreement of the Coulomb potential and the CRW data suggests that

the point-charge approximation of the CG sites can be justified. The energy components of the set $\{A_{vdW}, B_{ES}\}$ do not have these characteristic physical shapes. The electrostatic component is repulsive and the convergence of the attractive tail of the Van-der-Waals part does not converge to zero on short distances. In this case, the interpretation of the interaction components is not as straightforward and we could not justify the use of point charges from the physical form of the interaction potential.

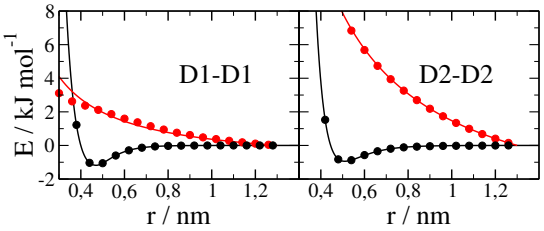


Figure 6: CG interaction potentials of the like pairs of DXE. The black circles represent the calculated B_{vdW} potential, while the black curve shows the fitted B_{vdW} potential. Red circles represent the calculated A_{ES} and the red curves show the theoretical prediction for a potential of point charges.

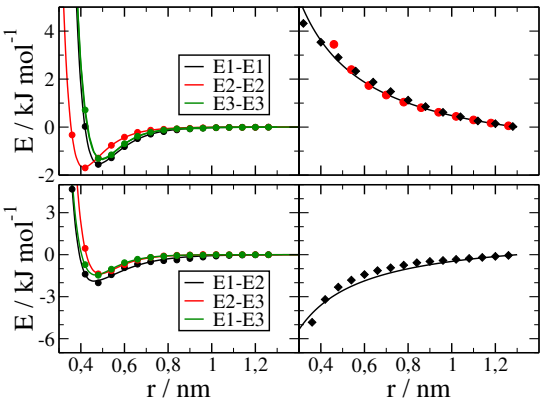


Figure 7: CG interaction potentials for all pairs in the EPP CG model. The curves in the left half show the fitted functions of B_{vdW}. Circles represent the calculated values of B_{vdW}. In the right half circles represent the calculated A_{ES} and lines show the theoretical prediction. Like pairs are shown in the upper half, unlike pairs in the lower half respectively. Note that the theoretically predicted potential is identical for the two like pairs and that for all pairs containing E3 $A_{ES} = 0$ (not shown here).

Table 1: Non-bonded parameters of the CG force field

	ε / kJ mol^{-1}	r_0 / nm	k_1 / nm^{-1}	k_2 / nm^{-1}	q / e
DXE					
D1-D1	1.182	0.471	9.739	13.583	-0.29
D2-D2	0.942	0.510	10.731	10.433	0.58
D1-D2	1.136	0.486	10.065	10.659	
EPP					
E1-E1	1.529	0.481	11.136	9.537	0.29
E2-E2	1.678	0.413	11.934	10.538	-0.29
E3-E3	1.315	0.495	10.550	10.841	0
E1-E2	1.887	0.458	10.828	7.348	
E2-E3	1.451	0.469	10.592	11.014	
E1-E3	1.362	0.489	10.867	10.920	

Figures 6 and 7 show the CG interaction potentials for the remaining pairs of DXE and EPP. Here we show the fitted Van-der-Waals component and the calculated CRW potential for the electrostatic component (symbols) as well as for the theoretical Coulomb interaction potential based on point charges (lines). Table 1 gives an overview of all the non-bonded parameters used in the CG force field.

Physical properties of the coarse-grained model systems

In this section, we present results from liquid phase simulations using the CRW-CG models described above and compare them to quantities obtained from simulations using the FG reference force field.

The properties investigated include the isobaric heat capacity c_p and the isothermal compressibility κ, which can be calculated from enthalpy and volume fluctuations, respectively, in the constant pressure-temperature ensemble:

$$c_p = \frac{\langle H^2 \rangle - \langle H \rangle^2}{k_B T^2 N_{DOF}} \tag{13}$$

$$\kappa = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B T \langle V \rangle} \tag{14}$$

where V denotes the total volume, T the temperature and k_B the Boltzmann constant. The total enthalpy $H = E_{kin} + E_{pot} + pV$ is calculated using the pressure p and the total kinetic (E_{kin}) and potential (E_{pot}) energies of the system. The number of thermal degrees of freedom $N_{DOF} = 3N_{at} - 3$, where N_{at} denotes the total number of atoms in the simulation.

In addition, we calculate the relative dielectric permittivity ε_r of the liquid using:[43]

$$\varepsilon_r = 1 + \frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3\varepsilon_0 V k_B T}, \tag{15}$$

where $\mathbf{M}$ is the total dipole moment of the system and ε_0 is the permittivity of the vacuum.

Another quantity relating to the electrostatic properties of the liquid is the finite system Kirkwood factor G_k which relates to the strength of correlation between molecular dipole moments and can be calculated as follows:[44]

$$G_k = \frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{N_{mol} \langle \mu^2 \rangle} \tag{16}$$

using the number of molecules N_{mol} and the molecular dipole moment μ.

From NVT simulations with a half empty box (extended to length L_z in z-direction) we calculate the surface tension from the pressure tensor components p_{ii}:

$$\gamma = \frac{L_z}{2} \left(p_{zz} - \frac{p_{xx} + p_{yy}}{2} \right) \tag{17}$$

Table 2: Properties at 300K

	DXE		EPP	
Property	FG	CG	FG	CG
ρ / kg m^{-3}	871	919	855	930
γ / mN m^{-1}	29.9 ± 0.4	26.2 ± 1.1	22.2 ± 0.4	26.8 ± 1.7
κ / 10^{-9} Pa^{-1}	0.97 ± 0.01	1.17 ± 0.02	1.18 ± 0.01	1.03 ± 0.04
c_p / J mol^{-1} K^{-1}	29.7 ± 0.1	16.7 ± 0.3	29.2 ± 0.1	11.5 ± 0.7
ε_r	10.44	30	4.12	10.7
G_k	1.15	2.8	1.05	1.25

Dimethoxyethane

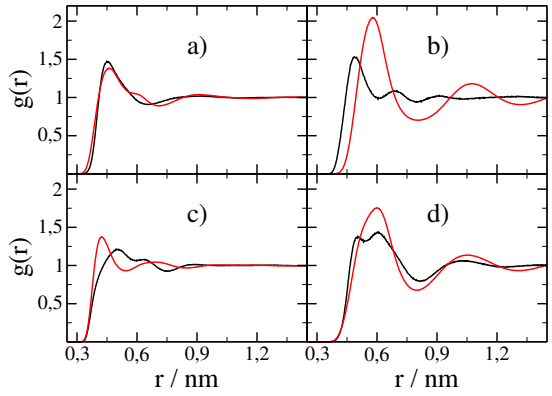


Figure 8: Radial distribution functions of the centers of mass of CG sites from simulations using the FG (black) and CG (red) models. a) D1-D1; b) D2-D2; c) D1-D2; d) RDF of molecular COM.

The RDF of all interaction pairs are shown in figure 8. While the RDF of the D1-D1 pair is almost quantitatively reproduced, the CG model yields only a rough qualitative reproduction for the remaining two pairs. Especially for the D2-D2 pair we can observe a loss of resolution and an increase in long-range correlations which is probably due to an increase in the effective excluded volume of the central methylene group. The CG model reproduces the excluded volume of the D1-D2 pair well but also fails to resolve details of the RDF of this pair qualitatively and quantitatively. The same can also be observed in the RDF of the molar centers of mass.

The values of the thermodynamic observables computed for the FG and CG models are shown in table 2 for a temperature of 300 K and in figure 9 for a range of temperatures.

The mass density of the liquid is overestimated by 5% at 300 K and the thermal expansion coefficient of the FG model is well reproduced by the CG model in the temperature range between 260 K and 340 K.

The heat capacity per degree of freedom predicted by the CG model is lower than the value calculated from FG reference simulations. This result is not surprising because the

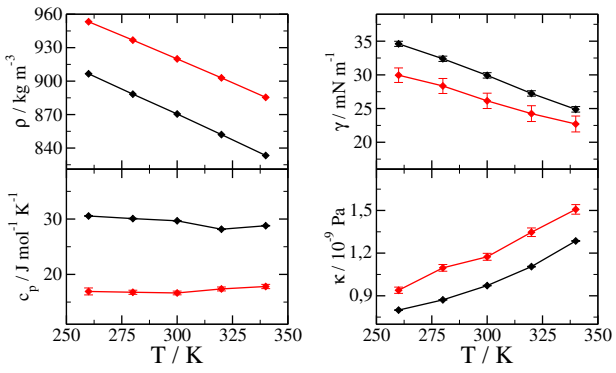


Figure 9: Calculated physical properties of liquid systems of DXE modeled with the FG (black) and CG (red) models as a function of temperature.

coarse-graining procedure softens the potential energy landscape of the system and therefore will reduce its capability to store thermal energy in its translational and internal degrees of freedom.

The isothermal compressibility is overestimated at all temperatures but the temperature dependence is reproduced by the CG model. We attribute this overestimation to an increase in volume fluctuations which are of higher magnitude in the CG model due to a softened potential energy surface. We would like to point out that a more accurate representation of the volume fluctuations of the system is in principle possible by the use of the barostat formulation for CG systems published recently by Dunn and Noid.[45] The use of this thermostat, however, requires additional parametrization that would go beyond the scope of this work.

The liquid-vapor surface tension for DXE is underestimated by $\sim 10\%$ and the temperature dependence is well reproduced in the given temperature range. These results document the excellent state point transferability of the CRW-CG models as reproduction of surface properties requires a very good description of both liquid and vapor behavior.

The value of ε_r is overestimated in the CG model although the molecular dipole moment distribution (shown in fig. 10, left panel) of the FG model is well reproduced by the CG model. A comparison of the Kirkwood factors shows that G_k is increased in the CG model

in comparison to the FG reference by approximately the same ratio as ε_r. This indicates that dipole-dipole correlations are too strong which may correlate with the overestimation of D2-D2 density correlations shown in fig. 8b.

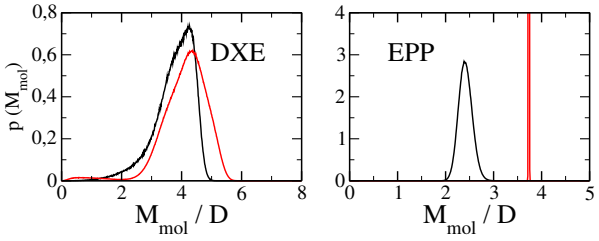


Figure 10: Distribution of molecular dipole moments in the FG (black) and CG (red) simulations of DXE and EPP.

Ethylpropionate

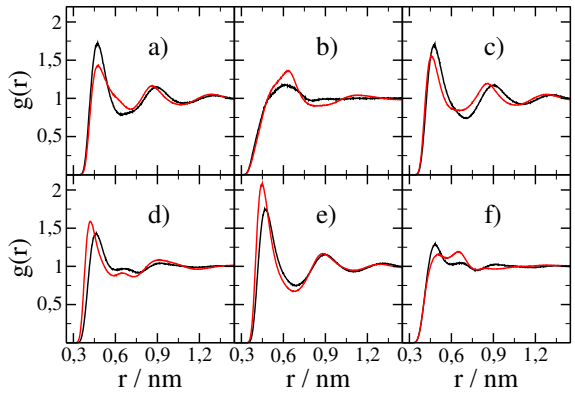


Figure 11: Radial distribution functions of the centers of mass of CG sites from simulations using the FG (black) and CG (red) models. a) E1-E1; b) E2-E2; c) E3-E3; d) E1-E2; e) E1-E3; f) E2-E3.

The $g(r)$ of all interaction pairs are shown in figure 11. The performance of the EPP CG model in reproducing the liquid structure is better than that of the DXE model. We observe a match of the RDF for all pairs and although a perfect quantitative reproduction

is not achieved (and should not be expected anyway) the CG model is able to reproduce all essential features of the reference liquid structure qualitatively.

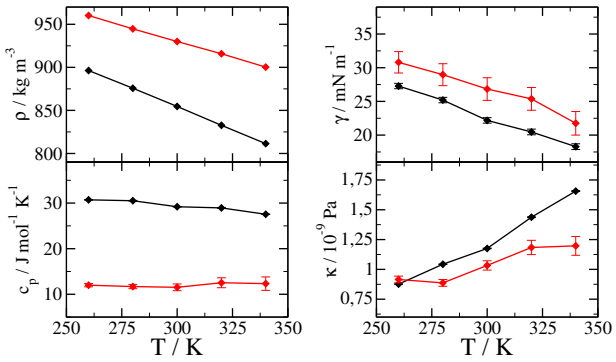


Figure 12: Calculated physical properties of liquid systems of EPP modeled with the FG (black) and CG (red) models as a function of temperature.

The mass density is slightly higher compared to the FG reference (overestimation of $\sim$ 9% at 300K) and the thermal expansion coefficient is less well matched by the CG model which predicts a lower expansion coefficient than the reference FG simulation.

The heat capacity is lower in the CG model, a finding that has already been discussed above for DXE.

The isothermal compressibility is underestimated for all temperatures and the temperature dependence is also different for the compared models. This finding is not intuitive because it might be expected that the softened potential energy surface in the CG model leads to increased volume fluctuations like it was observed for DXE. However we suspect the reason for the underestimation of κ is to be found in the higher density in the CG model which renders the liquid less compressible. The density is overestimated more strongly for EPP than for DXE and therefore density effects might trump the effect of the softer interactions. This hypothesis is supported by the fact that the mismatch in κ is larger at temperatures that show a larger mismatch in the liquid density.

The liquid-vapor surface tension prediction by the two investigated models differ by $\sim$ 15% and the temperature dependence is well reproduced.

The relative permittivity ε_r is overestimated in the CG simulations by a factor of 2.5 although the Kirkwood factor G_k is reproduced better than in the case of DXE. However we observe that the molecular dipole moment is overestimated in the CG model (fig. 10). This suggests that the reason for the mismatch in the relative permittivity is not so much due to increased orientational correlations between dipoles but rather due to the increase of the molecular dipole moment itself.

Conclusion

We have presented a modification to the CRW bottom-up coarse-graining method that allows for a separation of electrostatic and Van-der-Waals components in the CG model by stepwise introduction of the respective fine-grained interaction components. This procedure has been applied using two polar organic molecules as test cases. The effective pair potentials obtained can be decomposed into effective electrostatic and effective vdW interactions only when choosing the appropriate order of introducing the fine-grained interaction components. In simulations of the CG system, the new approach ascertains that the vdW part of the interaction free energy has converged at the cutoff. Furthermore, methods for long-range electrostatics can be used for a proper treatment of the electrostatic interactions of the model. The CRW-CG interaction potentials are derived from two fine-grained molecules in vacuo and are not parametrized to fit any structural or thermodynamic target quantity. Still, coarse-grained simulations reproduce liquid-state properties in satisfactory agreement with the fine-grained model. In particular, the bulk density and predict the thermal expansion coefficient are reproduced quite accurately. This shows that CRW models with explicit electrostatics show a high degree of transferability, as has been observed for CRW models of apolar molecules.[3,12,14] Furthermore, both CG models presented in this work predict the vapor-liquid surface tension in good agreement with reference simulations, an observation that can serve as further evidence for the state point transferability of CRW-CG potentials.

Unfortunately, the CRW models do not reproduce the relative dielectric permittivity of the FG systems. This quantity depends on a number of factors and it is unclear whether modifying one of these factors independently will result an a better reproduction of the dielectric properties. However, if a correct reproduction of this quantity is desired one could attempt to tune the charges on the CG sites in order to change the molecular dipole moment in systems where the orientational correlations between dipole moments are well reproduced (like EPP). However, we advice to perform such tuning carefully because whenever such a modification of the model is performed, the interaction free energy G_{eff} will no longer be independent of the introduction order and it is likely that the model will lose a great deal of its predictive power and transferability when such *a posteriori* modifications are applied.

We conclude that the modified CRW method presented in this work is a good extension to the toolbox of systematic bottom-up coarse-graining of polar molecular systems. The method can be especially useful for polymers or systems with a large number of different interaction sites. For these systems vacuum sampling is particularly advantageous as liquid state simulations will be computationally expensive and structure-based CG methods may be difficult to converge. Further studies remain to be performed to test the method on more polar molecules (e.g. alcohols, amines, ionic liquids) in order to determine, whether CG models can be obtained by our modified method also for these systems.

We finally point out that the effective-force coarse-graining method introduced by Wang and coworkers uses an approach that is methodologically very similar to the CRW method.[4] In their study, the authors also show that the CG interactions can be decomposed into effective Van-der-Waals and electrostatic contributions. As parametrization of EF-CG methods can be efficiently performed in the liquid phase differences can be expected with respect to the CRW potentials originating from vacuum sampling. Another methodological difference, however, lies within the fact that effective potentials derived with the CRW method contain indirect contributions originating from the modified sampling in a simulation with excluded interactions. In EF-CG, the sampling is only performed with all interactions present in the

system and therefore these contributions are absent. As a result of this subtle methodological difference we can expect slightly different potentials even if the fine-grained sampling is performed in an identical environment. The resulting CG potentials from the EF-CG and CRW methods will however show a similar shape, and it would be interesting to see to which degree the resulting models are comparable in terms of transferability and representability.

Acknowledgement

The authors thank Severine Rupp for her help in performing initial calculations for the EPP model. Financial support for this work was granted by the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center Transregio TRR 146 Multiscale Simulation Methods for Soft Matter Systems.

References

(1) Peter, C.; Kremer, K.; Multiscale simulation of soft matter systems. *Faraday Discuss.* **2010**, *144*, 9.

(2) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodríguez-Ropero, F.; Van der Vegt, N. F. A.; Systematic coarse-graining methods for soft matter simulations – a review. *Soft Matter* **2013**, *9*, 2108.

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

(4) Wang, Y.; Noid, W. G.; Liu, P.; Voth, G. A.; Effective force coarse-graining. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2002.

(5) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.;

Andersen, H. C.; The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **2008**, *128*, 244114.

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

(7) Lyubartsev, A.; Laaksonen, A.; Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **1995**, *52*, 3730–3737.

(8) Ganguly, P.; Mukherji, D.; Junghans, C.; Van der Vegt, N. F. A.; Kirkwood-Buff Coarse-Grained Force Fields for Aqueous Solutions. *J. Chem. Theory Comput.* **2012**, *8*, 1802–1807.

(9) Shell, M. S.; The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129*, 144108.

(10) Mullinax, J. W.; Noid, W. G.; A Generalized-Yvon–Born–Green Theory for Determining Coarse-Grained Interaction Potentials. *J. Phys. Chem. C* **2010**, *114*, 5661–5674.

(11) McCarty, J.; Clark, A. J.; Copperman, J.; Guenza, M. G.; An analytical coarse-graining method which preserves the free energy, structural correlations, and thermodynamic state of polymer melts from the atomistic to the mesoscale. *J. Chem. Phys.* **2014**, *140*, 204913.

(12) Brini, E.; Herbers, C. R.; Deichmann, G.; Van der Vegt, N. F. A.; Thermodynamic transferability of coarse-grained potentials for polymer–additive systems. *Phys. Chem. Chem. Phys.* **2012**, *14*, 11896.

(13) Fritz, D.; Herbers, C. R.; Kremer, K.; Van der Vegt, N. F. A.; Hierarchical modeling of polymer permeation. *Soft Matter* **2009**, *5*, 4556.

(14) Brini, E.; Van der Vegt, N. F. A.; Chemically transferable coarse-grained potentials from conditional reversible work calculations. *J. Chem. Phys.* **2012**, *137*, 154113.

(15) Wang, Y.; Feng, S.; Voth, G. A.; Transferable Coarse-Grained Models for Ionic Liquids. *J. Chem. Theory Comput.* **2009**, *5*, 1091–1098.

(16) Saunders, M. G.; Voth, G. A.; Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, *42*, 73–93.

(17) Cao, F.; Sun, H.; Transferability and Nonbond Functional Form of Coarse Grained Force Field – Tested on Linear Alkanes. *J. Chem. Theory Comput.* **2015**, *11*, 4760–4769.

(18) Hsu, D. D.; Xia, W.; Arturo, S. G.; Keten, S.; Systematic Method for Thermomechanically Consistent Coarse-Graining: A Universal Model for Methacrylate-Based Polymers. *J. Chem. Theory Comput.* **2014**, *10*, 2514–2527.

(19) 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.

(20) Qian, H.-J.; Carbone, P.; Chen, X.; Karimi-Varzaneh, H. A.; Liew, C. C.; Müller-Plathe, F.; Temperature-Transferable Coarse-Grained Potentials for Ethylbenzene, Polystyrene, and Their Mixtures. *Macromolecules* **2008**, *41*, 9919–9929.

(21) Mognetti, B. M.; Yelash, L.; Virnau, P.; Paul, W.; Binder, K.; Müller, M.; MacDowell, L. G.; Efficient prediction of thermodynamic properties of quadrupolar fluids from simulation of a coarse-grained model: The case of carbon dioxide. *J. Chem. Phys.* **2008**, *128*, 104501.

(22) Mognetti, B. M.; Virnau, P.; Yelash, L.; Paul, W.; Binder, K.; Müller, M.; MacDowell, L. G.; Coarse-grained models for fluids and their mixtures: Comparison of Monte

Carlo studies of their phase behavior with perturbation theory and experiment. *J. Chem. Phys.* **2009**, *130*, 044101.

(23) Mullinax, J. W.; Noid, W. G.; Extended ensemble approach for deriving transferable coarse-grained potentials. *J. Chem. Phys.* **2009**, *131*, 104110.

(24) Moore, T. C.; Iacovella, C. R.; McCabe, C.; Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *J. Chem. Phys.* **2014**, *140*, 224104.

(25) Hess, B.; Holm, C.; Van der Vegt, N.; Osmotic coefficients of atomistic NaCl (aq) force fields. *J. Chem. Phys.* **2006**, *124*, 164509.

(26) Villa, A.; Peter, C.; Van der Vegt, N. F. A.; Transferability of Nonbonded Interaction Potentials for Coarse-Grained Simulations: Benzene in Water. *J. Chem. Theory Comput.* **2010**, *6*, 2434–2444.

(27) Ardham, V. R.; Deichmann, G.; Van der Vegt, N. F. A.; Leroy, F.; Solid-liquid work of adhesion of coarse-grained models of n-hexane on graphene layers derived from the conditional reversible work method. *J. Chem. Phys.* **2015**, *143*, 243135.

(28) Fritz, D.; Harmandaris, V. A.; Kremer, K.; Van der Vegt, N. F. A.; Coarse-Grained Polymer Melts Based on Isolated Atomistic Chains: Simulation of Polystyrene of Different Tacticities. *Macromolecules* **2009**, *42*, 7579–7588.

(29) Dallavalle, M.; Van der Vegt, N. F. A.; Evaluation of mapping schemes for systematic coarse graining of higher alkanes. *Phys. Chem. Chem. Phys.* **2017**, *19*, 23034–23042.

(30) Horta, B. A. C.; Fuchs, P. F. J.; Van Gunsteren, W. F.; Hünenberger, P. H.; New Interaction Parameters for Oxygen Compounds in the GROMOS Force Field: Improved Pure-Liquid and Solvation Properties for Alcohols, Ethers, Aldehydes, Ketones, Carboxylic Acids, and Esters. *J. Chem. Theory Comput.* **2011**, *7*, 1016–1031.

(31) Plimpton, S.; Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

(32) Schneider, T.; Stoll, E.; Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **1978**, *17*, 1302–1322.

(33) Darden, T.; York, D.; Pedersen, L.; Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(34) Hills, R. D.; Lu, L.; Voth, G. A.; Multiscale Coarse-Graining of the Protein Energy Landscape. *PLoS Comput. Biol.* **2010**, *6*, e1000827.

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

(36) Hess, B.; Kutzner, C.; Van der Spoel, D.; Lindahl, E.; GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(37) 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.

(38) Nosé, S.; A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(39) Hoover, W.; Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

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

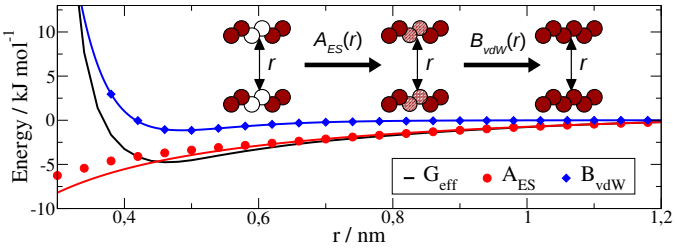
(41) Nosé, S.; Klein, M.; Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

(42) 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.

(43) Neumann, M.; Computer simulation and the dielectric constant at finite wavelength. *Mol. Phys.* **1986**, *57*, 97–121.

(44) Glättli, A.; Daura, X.; Van Gunsteren, W. F.; Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L. *J. Chem. Phys.* **2002**, *116*, 9811–9828.

(45) Dunn, N. J. H.; Noid, W. G.; Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *J. Chem. Phys.* **2015**, *143*, 243148.


Energy / kJ mol-1
10
5
0
-5
-10
0,4
0,6
0,8
1
1,2
r / nm
r
A_ES(r)
B_vdW(r)
G_eff
A_ES
B_vdW