Transferability of local density assisted implicit solvation models for homogeneous fluid mixtures

David Rosenberger,*,† Tanmoy Sanyal,‡ M. Scott Shell,*,‡ and Nico F. A. van der Vegt†

†*Eduard Zintl Institut für Anorganische und Physikalische Chemie, Technische Universität Darmstadt, Darmstadt, Germany*

‡*Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, California, United States*

E-mail: rosenberger@cpc.tu-darmstadt.de; shell@engineering.ucsb.edu

Abstract

The application of bottom-up coarse grained (CG) models to study the equilibrium mixing behavior of liquids is rather challenging, since these models can be significantly influenced by the density or the concentration of the state chosen during parametrization. This dependency leads to low transferability in density/concentration space and has been one of the major limitations in bottom-up coarse graining. Recent approaches proposed to tackle this shortcoming range from the addition of thermodynamic constraints, to an extended ensemble parametrization, to the addition of supplementary terms to the system's Hamiltonian. To study fluid phase equilibria with bottom-up CG models, the application of local density (LD) potentials appears a promising approach, as shown in previous work by Sanyal and Shell [T. Sanyal, M. S. Shell, J. Phys. Chem. B, 2018, 122, 5678]. Here, we want to further explore this method and test its ability to model a system which contains structural inhomogeneities only on the molecular

scale, namely solutions of methanol and water. We find that a water-water LD potential improves the transferability of an implicit-methanol CG model towards high water concentration. Conversely, a methanol-methanol LD potential does not significantly improve the transferability of an implicit-water CG model towards high methanol concentration. These differences appear due to the presence of cooperative interactions in water at high concentrations that the LD potentials can capture. In addition, we compare two different approaches to derive our CG models, namely, relative entropy optimization and the Inverse Monte Carlo method, and formally demonstrate that the analytical and numerical assumptions under these two methods yield equivalent results.

1 Introduction

Our understanding of the driving forces behind processes in soft condensed matter has greatly benefited from computer simulations, and from molecular dynamics (MD) simulations in particular. This technique allows an atomistic view into complex systems, but is limited by computational overhead to modeling length and time scales of tens of nanometers and microseconds, respectively. This hurdle can be overcome with coarse-grained (CG) particle models whose number of degrees of freedom (DOFs) and interactions to be evaluated are significantly smaller compared with fine-grained (FG) models. Apart from their computational advantages in computer simulations, CG models, and methods to derive them, provide important additional merits. Efforts to make models as simple as possible provide additional insight into emergent driving forces that may not be as easily obtained based on FG models alone. Systematic removal of DOFs involved in deriving a CG model based on its FG counterpart is the basis for bottom-up or systematic coarse graining.[1–3] To reduce DOFs, the high resolution, or FG, configuration space is projected onto a CG configuration space lower in resolution. In order to evaluate the corresponding free energy surface, the so

called multibody potential of mean force (PMF) must be calculated[2]

$$W(\mathbf{R}) = -k_B T \ln \int_V d\mathbf{r}\ \exp[-\beta U_{FG}(\mathbf{r})]\ \delta(\mathbf{R} - \mathbf{M}(\mathbf{r})) \tag{1}$$

where $W(\mathbf{R})$ is the multibody PMF and $\beta = 1/k_B T$ with k_B the Boltzmann constant and T the temperature. $\mathbf{M}$ is the projection, or mapping, operator, which relates a FG configuration $\mathbf{r}$ to a CG one $\mathbf{R}$. Due to the highly multibody nature of the integral, $W(\mathbf{R})$ is too complicated to compute exactly in practice, and only approximate solutions are possible. Solving an inverse problem is one way to obtain such an approximate solution to the multibody PMF. The goal is then to find a CG model that accurately reproduces one or multiple quantities of the FG model in the CG configuration space, by minimizing the difference between the FG and CG configuration space with respect to the quantity chosen. This inverse problem can be solved by means of either a variational principle[4–9] or by application of iterative Newton or quasi-Newton inversion techniques.[10–13]

Despite the methodological differences, all bottom-up coarse grained models suffer from the same two fundamental problems: accurate representability and state point transferability. The representability characterizes the ability of a CG model to simultaneously reproduce multiple properties of the FG system like structure, pressure and isothermal compressibility. The transferability instead describes the applicability of CG models at state points not included in the parametrization. Whereas accurate representability is guaranteed for the target property chosen to solve the inverse problem, it is not guaranteed for other properties, i.e. a match in structure does not automatically guarantee a match in thermodynamics or dynamics, or a match in forces does not guarantee structural agreement.[14–18] An equally outstanding challenge is the transferability of CG models.[19–22] The main reason why it is difficult to achieve both is that entropic contributions from particles "lost" upon coarse graining are missing. Therefore, the generated effective pair potentials cannot capture changes in entropy necessary to describe certain thermodynamic properties and necessary to be trans-

ferable to different state points.[16,23,24] Several approaches have been proposed to tackle both representability and transferability. Among the most common are the addition of thermodynamic constraints to account for accurate pressure,[12] Kirkwood-Buff integrals[25] or the surface tension,[26,27] the application of an extended ensemble parametrization,[28,29] or the use of a extended Hamiltonian description for the energy of the system.[30–35]

Allen and Rutledge proposed the idea of local density (LD) dependent interactions to improve implicit solvent models.[20] Interactions based on the LD of CG sites have been used to improve the transferability[36] and the representability[37] of CG models. The local density of a CG site is simply a weighted local co-ordination number around that site, and can be written down generally for arbitrary combinations of central and neighboring types of sites. The LD potential can then be cast as a function of the local density and added as a corrective extension to the traditional pair-wise form of the Hamiltonian in CG models. Unlike pairwise Hamiltonians, a LD potential incorporates information about the inherently multibody environment around CG sites which contributes to enhanced model transferability. Further, it is a *mean-field* potential with computational complexity similar to that of pair potentials and thus does not sacrifice computational speed (further implementation details can be found in Ref [32]). LD potentials have been used to improve the sampling of conformation space in implicit solvent models of superhydrophobic polymers and were found to enhance the model's transferability to different polymer lengths.[32] Recently, they have also been shown to improve structural transferability in CG models of liquid mixtures such as benzene in water.[34]

In this work we want to test the ability of LD dependent CG models to quantitatively and qualitatively describe mixtures of water and methanol. These mixtures provide an interesting test case for the LD dependent potentials, since they show strong microheterogeneities at atomistic length scales as a function of methanol concentration, while remaining miscible at a macroscopic scale.[38–43]Laaksonen et al. showed that these microheterogeneities are caused by a non-homogeneous distribution of the two components in the mixture.[44] This is expressed

through structural patterns determined by the dominant component in the system. X-ray emission spectroscopy experiments applied by Guo et al. revealed that the inhomogeneous mixing between methanol and water can be explained by the formation of rings of methanol bridged by water molecules.[41] In agreement, Perera et al. found methanol molecules forming chain-like structures, caused by water bridging the hydroxyl groups of the methanol molecules.[45] Further, Pascal and Goddard confirmed the picture of incomplete mixing.[46] At low methanol concentrations, methanol molecules bury their hydrophobic groups away from water. With increasing methanol concentration this is no longer possible and free mixing is observed. At high methanol concentrations, the system is best described as water dissolved in methanol.[46]

Classical MD simulations of methanol-water mixtures have almost exclusively been based on all-atom force field models.[38,44–46] It is, however, interesting to ask if CG particle models can equally well be used to describe the structural properties of these systems. This question may provide insights to modeling large-scale phenomena driven by the interplay between hydrophobic and hydrophilic interactions in complex systems not amenable to all-atom models. In this study, we investigate the possibility to study this interplay with simple CG models. We do this by explicitly accounting for LD effects and examine if LD potentials can effectively describe the micro-heterogeneities observed in water/methanol mixture with a simple single site CG model for both liquids, water and methanol in an implicit solvent environment. We derive two different CG models: (I) CG methanol in implicit water and (II) CG water in implicit methanol. Further, we compare two different methods to generate bottom up CG models, namely Inverse Monte-Carlo (IMC)[11] and relative entropy optimization,[13] and we show analytically and numerically under which assumptions these two methods are equivalent.

The remainder of the article is structured as follows: first, the basic theoretical background on IMC and relative entropy optimization is given. Second, we prove analytically under which assumptions IMC and the relative entropy method are equivalent. Next, we briefly

discuss the extension of the relative entropy method to LD potentials, followed by the details of the numerical calculations performed. Adjacent, we present the main results of this study along with a detailed discussion, followed by the conclusion.

2 Methods

2.1 Inverse Monte Carlo

The Inverse Monte Carlo (IMC) Method, or Newton inversion method (introduced by Lyubartsev and Laaksonen) aims to derive a CG force field (FF) that reproduces the pairwise structure, i.e. the radial distribution function (RDF), of the underlying atomistic or fine grained (FG) system.[11] The CG FF is estimated initially as the two-body potential of mean force (PMF, $U^0(r_{ij})$) acting between two particles i and j along the distance r_{ij} obtained from the corresponding RDF of the FG system ($g^0(r_{ij})$):

$$U^0(r_{ij}) = -k_B T \ln g^0(r_{ij}) \tag{2}$$

In many cases, the PMF does not accurately resemble the effective pair potential in the CG configuration space due to the relevance of higher order correlations. Thus, the potential is updated a series of times n, solving a set of linear equations until the difference in the RDF is minimized. This leads to numerical pair potentials.

The set of linear equations is given by Eq.(3), where N_α is the number of particle pairs at a distance α either in the CG system or in the mapped reference system (N_α^0), $\boldsymbol{J}$ is a Jacobian matrix, ΔU is the potential update and γ is a particle pair distance $> \alpha$. The Jacobian matrix is defined in Eq.(4), where N_γ is the number of particle pairs separated by a distance

γ. Finally, the update of the PMF gets computed according to Eq.(5).

$$\langle N_\alpha \rangle^{CG} - N_\alpha^0 = \boldsymbol{J}_{\alpha,\gamma} \Delta U_\gamma \tag{3}$$

$$\boldsymbol{J} = \frac{\partial \langle N_\alpha \rangle^{CG}}{\partial U_\gamma} = -\beta(\langle N_\alpha N_\gamma \rangle - \langle N_\alpha \rangle \cdot \langle N_\gamma \rangle) \tag{4}$$

$$U^n(r_{ij}) = U^{n-1}(r_{ij}) - \boldsymbol{J}^{-1}(g^{n-1}(r_{ij}) - g^0(r_{ij})) \tag{5}$$

The number of particle pairs N_α is related to the g(r) via[47]

$$\langle N_\alpha \rangle = \frac{N(N-1)}{2} \frac{4\pi r_\alpha^2 \Delta r}{V} g(r_\alpha) \tag{6}$$

where V is the volume of the system, N the number of particles and Δr is the discretization grid spacing. The Jacobian defined in Eq.(4) explicitly contains cross correlations between the number of particle pairs at different distances. This provides several advantages in converging towards the final effective pair potentials, and some disadvantages in terms of numerical stability of IMC, compared to the similar Iterative Boltzmann Inversion method[10,12] as discussed in the literature.[26,48,49]

2.2 Relative entropy optimization for pair potentials

Another way to determine a CG FF is through relative entropy optimization, as proposed by Shell.[13] The relative entropy is a quantity that measures the information loss upon reducing the FG system to the CG model. It is defined as

$$S_{\text{rel}} = \sum_i p_0(i) \ln \frac{p_0(i)}{p_{CG}(i)} + S_{\text{map}} \tag{7}$$

where p is the probability to observe a certain configuration i determined either by the FG FF (p_0) or by the CG FF (p_{CG}) and S_{map} is a mapping entropy which accounts for the degeneracy of atomistic states in the CG configuration space. In the canonical ensemble the

relative entropy can be expressed as

$$S_{\text{rel}} = \beta\langle U_{CG}(\boldsymbol{\zeta}) - U_{FG}\rangle_{FG} - \beta(A_{CG}(\boldsymbol{\zeta}) - A_{FG}) + S_{\text{map}} \tag{8}$$

where U is the potential energy, A is the Helmholtz free energy and $\boldsymbol{\zeta}$ is a vector that contains all parameters of the CG FF. The idea in relative entropy optimization is to find the optimal set of parameters $\boldsymbol{\zeta}$ that minimize the information loss between the FG and CG system. Similar to IMC, where the pair potential U is updated, in relative entropy optimization the FF parameters $\boldsymbol{\zeta}$ are updated until a minimum in the relative entropy is reached.[50] Using a Newton-Raphson approach, the parameter update scheme is:

$$\boldsymbol{\zeta}^k = \boldsymbol{\zeta}^{k-1} - \boldsymbol{H}^{-1}\nabla S_{\text{rel}} \tag{9}$$

In Eq.(9), $\boldsymbol{H}$ is the Hessian matrix and ∇S_{rel} is the gradient of the relative entropy. Sequential iterations successively bring the parameters to a local S_{rel} minimum. This scheme is only applicable if the Hessian is positive definite, otherwise a steepest descent or conjugate gradient optimization scheme is applied.[50,51] It is interesting to note that if U consists of splines or tabulated potentials (as in IMC), then the relative entropy has a single global minimum and Eq.(9) always applies.[51]

2.3 Equivalence between Inverse Monte Carlo and relative entropy optimization

The two methods, IMC and relative entropy optimization (when using a Newton-Raphson update scheme), are equivalent for the derivation of effective pair potentials if the following

applies:

$$\nabla S_{\mathrm{rel}} = (\langle N_\alpha \rangle^{CG} - N_\alpha^0)) \tag{10}$$

$$\boldsymbol{H}^{-1} = \boldsymbol{J}^{-1} \tag{11}$$

For simplicity we assume a system with only one component, but the equations can be easily extended to multi component systems as well.

As proposed by Lyubartsev and Laaksonen,[11] we start with a discretized Hamiltonian (U) to describe the potential energy of the system

$$U = \sum_\alpha N_\alpha(q)\zeta_\alpha \tag{12}$$

where ζ_α is the pair potential and N_α is the exact number of particle pairs at distance α, given by:

$$\langle N_\alpha \rangle = \frac{\int dq N_\alpha \prod_\lambda \exp(-\beta N_\lambda(q)\zeta_\lambda)}{\int dq \prod_\lambda \exp(-\beta N_\lambda(q)\zeta_\lambda)} \tag{13}$$

The Jacobian in Eq.(4) is then given by

$$\frac{\partial \langle N_\alpha \rangle}{\partial \zeta_\gamma} = \frac{\partial}{\partial \zeta_\gamma} \left(\int dq N_\alpha \prod_\lambda \exp(-\beta N_\lambda(q)\zeta_\lambda) \cdot \frac{1}{\int dq \prod_\lambda \exp(-\beta N_\lambda(q)\zeta_\lambda)} \right) \tag{14}$$

which by application of the chain rule results in:

$$\boldsymbol{J} = \frac{\partial \langle N_\alpha \rangle}{\partial \zeta_\gamma} = -\beta(\langle N_\alpha N_\gamma \rangle - \langle N_\alpha \rangle \cdot \langle N_\gamma \rangle) \tag{15}$$

By inserting Eq.(15) in Eq.(3), the potential update, ΔU_γ, in IMC is computed by solving:

$$\langle N_\alpha \rangle^{CG} - N_\alpha^0 = (-\beta(\langle N_\alpha N_\gamma \rangle - \langle N_\alpha \rangle \cdot \langle N_\gamma \rangle))\Delta U_\gamma \tag{16}$$

where we define the left hand side (l.h.s) as

$$\Delta N_\alpha^{IMC} \equiv \langle N_\alpha \rangle^{CG} - N_\alpha^0 \tag{17}$$

Next, we evaluate the relative entropy in Eq.(8). Its derivative can be written as:

$$\nabla S_{\text{rel}} = \beta \left\langle \frac{\partial U}{\partial \zeta} \right\rangle_{FG} - \beta \left\langle \frac{\partial U}{\partial \zeta} \right\rangle_{CG} \tag{18}$$

Substituting U defined by Eq.(12), we obtain:

$$\nabla S_{\text{rel}} = N_\alpha^0 - \langle N_\alpha \rangle^{CG} \equiv \Delta N_\alpha^{S_{\text{rel}}} \tag{19}$$

In comparison with Eq.(17), the result of Eq.(19) reveals the following relation:

$$-\Delta N_\alpha^{S_{\text{rel}}} = \Delta N_\alpha^{IMC} \tag{20}$$

The Hessian matrix in Eq.(9) is given by

$$\boldsymbol{H} = \left\langle \frac{\partial^2 U}{\partial \zeta_\alpha \partial \zeta_\gamma} \right\rangle_{FG} - \left\langle \frac{\partial^2 U}{\partial \zeta_\alpha \partial \zeta_\gamma} \right\rangle_{CG} + \beta \left\langle \frac{\partial U}{\partial \zeta_\alpha} \frac{\partial U}{\partial \zeta_\gamma} \right\rangle_{CG} - \beta \left\langle \frac{\partial U}{\partial \zeta_\alpha} \right\rangle_{CG} \left\langle \frac{\partial U}{\partial \zeta_\gamma} \right\rangle_{CG} \tag{21}$$

where the first two terms vanish due to the linearity of U in the parameters ζ.[51] The linearization that is exploited here results from using cubic splines for the effective pair potentials in the relative entropy method. The fact that the CG pair potential is linear in its parameters (spline knots) is crucial to the success of a simple scheme like Newton-Raphson descent in discovering a global minimum on the relative entropy surface which then theoretically guarantees robust representability at least for the pair correlations. Using the Hamiltonian

defined in Eq.(12), the remainder is:

$$\boldsymbol{H} = \beta\langle N_\alpha N_\gamma\rangle - \beta\langle N_\alpha\rangle\langle N_\gamma\rangle \tag{22}$$

which leads to a similar relation as in Eq.(20)

$$\boldsymbol{H}^{-1} = -\boldsymbol{J}^{-1} \tag{23}$$

Taking these results and applying the definition for the potential update (see Eq.(5) and Eq.(9)) we end with the following relation:

$$\boldsymbol{H}^{-1}(-\Delta N_\alpha^{S_{rel}}) = -\boldsymbol{J}^{-1}\Delta N_\alpha^{IMC} \tag{24}$$

$$\boldsymbol{H}^{-1}\Delta N_\alpha^{S_{rel}} = \boldsymbol{J}^{-1}\Delta N_\alpha^{IMC} \tag{25}$$

This proves that under the assumption of a discretized Hamiltonian (tabulated in IMC and represented with splines in the relative entropy method), quasi-Newton optimization strategies like Newton-Raphson leads to an exact equivalence of the pair potentials obtained through IMC and relative entropy. Moreover this shows that relative entropy optimization leads to a match in the RDF between the FG and CG system without directly using it as a target quantity in the optimization process. It is important to note that the agreement between IMC and relative entropy minimization is likely to also hold when very knot-dense spline potentials are employed in the latter (instead of discretized ones), although the necessary knot density and discretization to observe quantitative agreement may be high.

2.4 Relative entropy optimization for local density potentials

Recently Sanyal and Shell applied the relative entropy approach not only for pair potentials, but also for so called local density (LD) potentials.[32,34] LD potentials (U_{LD}) account for the effect of neighboring particles on the effective pair potential (u_{pair}). This additional

contribution changes the total CG potential energy to:

$$U_{CG} = \sum_{i<j} u_{\text{pair}} + U_{LD} \tag{26}$$

U_{LD} is a sum over an unspecified function of the local density around each particle i in the system,

$$U_{LD} = \sum_{i} f(\rho_i) \tag{27}$$

where $f(\rho_i)$ is practically represented using cubic B-splines. Here, the local density, ρ_i, is the total number of neighboring particles within a specified and smoothed cutoff (r_c)

$$\rho_i = \sum_{j \neq i} \phi(r_{ij}) \tag{28}$$

where the indicator function (ϕ) adopts a value of 1 below an inner cutoff r_0 but continuously and quickly decays to 0 at r_c. The shape of this function is chosen to be computationally convenient and does not require the calculation of the absolute distance between pairs of particles (or any square root operations).

$$\phi(r) = \begin{cases} 1 & r \leq r_0 \\ c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 & r \in (r_0, r_c) \\ 0 & r \geq r_c \end{cases} \tag{29}$$

The difference between r_0 and r_c will be called Δ and is of the order of $0.1 - 0.12$ nm. The exact form of the coefficients c as well as more details on the LD potentials can be found in the original work by Sanyal and Shell and the recent extension to binary mixtures.[32,34]

3 Simulation Details

FG and IMC CG simulations are performed with the Gromacs-5.1.2 MD engine.[52,53] IMC potentials are generated with the VOTCA coarse graining package (version 1.4).[47,54] The relative entropy optimization is achieved with an in house code. To account for local density potentials, CG simulations are executed with a modified version of the LAMMPS simulation package that includes a custom local density potential.[55]

3.1 Fine grained simulations

All systems studied contain 5000 molecules in total, with different methanol mole fractions, x_M, of interest, namely 0.1, 0.5 and 0.9 in SPC/E water.[56,57] For FG methanol a Kirkwood-Buff based force field for united atoms is used.[58] Newton's equations of motion are integrated based on a leap-frog algorithm with a timestep of 1 fs. All systems are equilibrated for 2 ns at a constant pressure of 1 bar and at a constant temperature of 300 K (NPT condition). For both the barostat and thermostat, the weak coupling method of Berendsen is applied with a coupling constant of $\tau_p = 1$ ps for the barostat and $\tau_T = 0.5$ ps for the thermostat.[59] The barostat compressibility is $4.5 \cdot 10^{-5}$ bar^{-1}. For the short-range van der Waals interactions a cut-off of 1.2 nm is applied with a long-range dispersion correction. Electrostatic interactions are treated with the particle-mesh-Ewald method[60] with a real space cut-off of 1.2 nm and a grid size of 0.12 nm. Bonded interactions in methanol are constrained with the LINCS algorithm.[61,62] Periodic boundary conditions are applied in the x, y, and z directions. The short equilibration is followed by a 10 ns run under NPT conditions replacing the Berendsen barostat and thermostat with the Parrinello-Rahman barostat[63] and the Nosé-Hoover thermostat[64] respectively. The coupling constants are set to $\tau_p = 1$ ps for the barostat and $\tau_T = 0.5$ ps thermostat. All other parameters are kept the same during the short equilibration run. The average volume of the 10 ns NPT simulation is then used

for the final production run under constant volume (NVT) conditions. Besides the volume constraint, all other parameters are the same as during the 10 ns NPT run.

3.2 Implicit solvent model

To generate the CG model for the united atom model of methanol in implicit SPC/E water[56,57] and for SPC/E water[56,57] in implicit methanol, a 3 to 1 mapping scheme is applied, where each molecule is mapped to its center of mass as illustrated in figure 1.

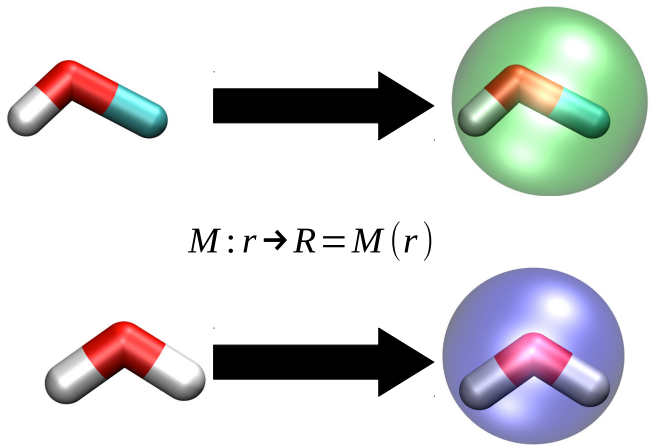


Figure 1: top: mapping scheme for CG methanol with the FG model on the left and the 1 bead representation (green) on the right; bottom: mapping scheme for CG water with the FG model on the left and the 1 bead representation (blue) on the right.

3.2.1 Inverse Monte Carlo optimization

On the basis of the FG RDFs between the centers of mass of water and water (WW) and methanol and methanol (MM) at $x_M = 0.5$, implicit solvent models for CG water and CG methanol are derived with the IMC method. As an initial guess for the iterative procedure, the PMF is taken (see Eq. (2)). To generate the CG configuration of the atomistic model,

water and methanol are mapped as illustrated in figure 1 and depending on the case – either water in implicit methanol or methanol in implicit water – the second component is implicitly present. This results in an uncharged single-site representation of the former 3 atom molecules. The effect of the lost component (either water in CG methanol or methanol in CG water) as well as the electrostatic interactions are integrated into the effective pair interaction generated via IMC. For the CG models of water in implicit methanol the RDFs are evaluated between 0.24 and 1.2 nm with a grid spacing of 0.01 nm. For CG methanol in implicit water the interval is changed to 0.3 and 1.5 nm with the same grid spacing. At each iteration a short MD simulation of 2 ns is performed with a timestep of 1 fs and with a leap-frog stochastic dynamics integrator. The simulations are performed at 300 K ((τ_T = 1.0 ps)) under NVT conditions using the average volume of the atomistic simulations at a mole fraction of 50% methanol ($x_M = 0.5$). No long-range dispersion correction for energy or pressure is applied for the numerical potentials during the CG simulation. The cut-off for the van der Waals interactions are the same as the maximum distance in the RDF evaluation. The iterative procedure is run until no change in the RDFs as well as in the potentials is further observed.

3.2.2 Relative entropy optimization

Relative entropy optimization is performed for the same CG models as described for IMC. The effects of the second component and the charges are again built into the effective interactions. The non-bonded pair potentials and the local density potentials are represented by cubic splines whose knot point values are determined during the optimization process. For both interactions 80 knot points are optimized. Models with only non-bonded pair interactions will be referred to as REO (Relative Entropy Optimization) models, whereas the case with pair splines and a local density potential will be called REO LD. The relative entropy at each iteration is calculated from short trial MD simulations launched with current estimates of the REO potentials. REO makes use of a reweighting scheme to decrease the number of trial MD runs thus reducing overall computation cost and statistical error.[50] During the

trial MD simulations the following steps are executed. First, the system is energy minimized with a conjugate gradient algorithm for 1000 steps, followed by 1 ns equilibration. Second, a 2 ns production run is performed under NVT conditions at 300 K. Newton's equations of motion are integrated according to the velocity-Verlet algorithm with a timestep of 1 fs. A Langevin thermostat is applied with a coupling constant of 0.1 ps The cut-off value for the REO potentials as well as the values for the outer cut-off (r_c in Eq. 29) and the difference between the outer and inner cut-off of the LD potentials (Δ) are listed in table 1.

Table 1: Cut-off values for the pair-potentials (r_c) and the outer cut-off values (r_c) for the LD potentials together with the difference between inner and outer cut-off for the LD potentials (Δ).

interaction	type	r_c(nm)	Δ (nm)
Methanol-Methanol (MM)	REO pair potentials	1.5	
	REO LD potentials	0.63	0.1
Water-Water (WW)	REO pair potentials	1.2	
	REO LD potentials	0.34	0.1

3.2.3 Coarse grained simulations

On the basis of the derived CG models, MD simulations are performed for 10 ns under NVT conditions at three different mole fractions of methanol $x_M = 0.1$, $x_M = 0.5$ and $x_M = 0.9$. The simulation parameters are the same as the ones used for the MD production phases in the iterative optimization approaches for IMC and relative entropy optimization.

4 Results and Discussion

4.1 Kirkwood-Buff Analysis of the fine grained system

In order to compare the derived CG models with their parent reference systems beyond structural accuracy, we also compute Kirkwood-Buff integrals (KBIs). KBIs relate local structure and thermodynamic properties like activity coefficients, solvation free energies or

the isothermal compressibility of stable mixtures.[65] For mixture components i and j they are defined as:[66]

$$G_{ij} = 4\pi \int_0^\infty [g_{ij}(s) - 1] s^2 ds \tag{30}$$

KBIs can be interpreted as the excess coordination number of particles j around a central particle i. This means that the larger the KBI value, the higher is the affinity between particles i and j and the more compressible is the system. To evaluate Eq.(30) in computer simulations, a thermodynamic limit needs to be taken while the system should be open with respect to its components. Notwithstanding the latter requirement, KBIs can be calculated in computer simulations of closed (NVT or NPT) systems. To this end, the integral in Eq.(30) is usually truncated at an upper integration limit $r < L/2$ (with L the linear simulation box dimension) where $G_{ij}(r)$ (the running KBI or RKBI) is observed to oscillate around a mean, plateau value. In recent work, it has been shown that this mean value corresponds to the thermodynamic limiting value of Eq. (30). A more detailed discussion on the issue of finite size and ensemble effects is beyond the scope of this work and can be found elsewhere.[67–69]

By applying Eq.(30) one critical problem occurs that should not be ignored: RDFs do not strictly approach a limiting value of 1 in closed systems. This leads to a drift in the asymptotic behavior of the RKBIs. The drift is caused by depletion or accumulation of particles j around a particle i at local scales. This local depletion or accumulation is then compensated by a positive or negative excess of particles j at long distances, since the total number of particles j is constant. This leads to incorrect limiting behavior ($r \to \infty$) of $g_{ij}(r)$, which must be corrected. We here use the empirical correction introduced by Ganguly and van der Vegt[22]

$$g_{ij}^{corr}(r) = g_{ij}(r) \frac{N_j(1 - \frac{(4/3)\pi r^3}{V})}{N_j(1 - \frac{(4/3)\pi r^3}{V}) - \Delta N_{ij}(r) - \delta_{ij}} \tag{31}$$

where N_j is the number of particles j, V is the volume of the system, ΔN_{ij} is the excess number of particles j around a particle i within a sphere of radius r and δ_{ij} is the Kronecker

delta. KBIs obtained from $g_{ij}^{corr}(r)$ will be named Ganguly (GKBI) in the following.

Figure 2 shows the RKBIs of the FG simulations at $x_M = 0.5$. One clearly sees a drift in the asymptotic behavior of the water-water (WW) RKBI (solid black line) and how the empirical correction of Ganguly (solid red line) shifts the RKBI to larger values at longer distances. This is also observed for methanol-methanol (MM) (dotted lines). In case of methanol-water (MW, dashed lines), one observes that the RKBIs are already well converged without any correction. Here the empirical correction of Ganguly introduces a small shift in the tails of the KBIs towards lower values. Now if one compares the RKBI values averaged between 1.0 and 1.5 nm, the influence of the Ganguly correction is fairly small and the effect of the empirical correction becomes important only at larger distances.

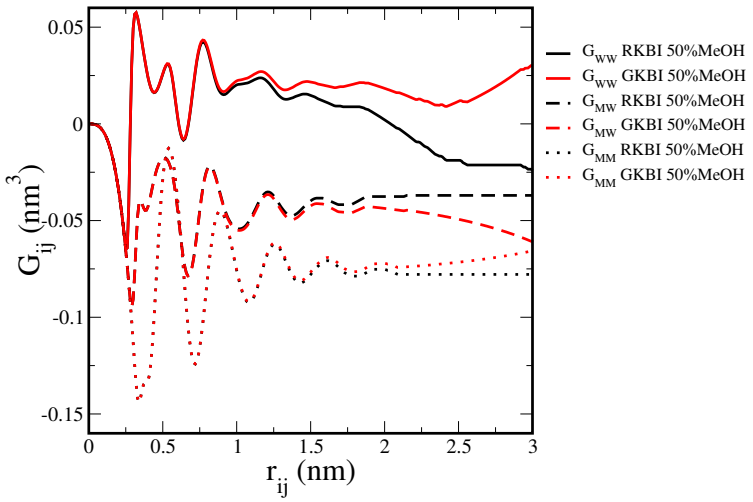


Figure 2: Influence of the Ganguly correction on the convergence of the Kirkwood-Buff Integrals in the AA system for Water-Water (solid lines), Methanol-Water (dashed lines) and Methanol-Methanol (dotted lines) at a mole fraction of x_M=0.5:uncorrected RKBI (black) and the RDF correction of Ganguly, GKBI, (red).

The average values for all KBIs obtained from FG simulations are presented in table 2. The averages are taken between 1.0 and 1.5 nm and the errors calculated by averaging over 5 independent simulations according to

$$err = \sqrt{\frac{\sum_i^N (G_{ij} - \langle G_{ij} \rangle)^2}{N(N-1)}} \tag{32}$$

where $N = 5$ is the number of simulations. As one sees, all KBI values are nearly the same no matter if the Ganguly correction is applied or not. The largest discrepancy between the RKBI and GKBI occurs for WW at $x_M = 0.9$. Further, one observes a minimum in G_{MM} at $x_M = 0.1$, which means that at low methanol concentrations the methanol molecules are mutually stronger depleted than at higher concentrations. This is in agreement with the work of Laaksonen et al.,[44] who showed that at low x_M the methanol molecules are separated by larger distances. Going from $x_M = 0.1$ to $x_M = 0.5$, G_{WW} and G_{MM} become larger (less negative). G_{WW} even changes sign and becomes positive. This indicates stronger water-water and methanol-methanol association in the $x_M = 0.5$ mixture. Upon further increasing the methanol content from $x_M = 0.5$ to $x_M = 0.9$, the values of G_{MW} in turn become larger while G_{WW} becomes smaller. Thus, at high mole fractions of methanol, the water molecules preferentially interact with methanol molecules. This observation agrees with the work of Pascal and Goddard.[46] The increase in G_{WW} upon raising the methanol concentration is also supported by other studies, which found that small water aggregates are formed the more methanol molecules are present in the system.[44–46] We note that the values of G_{MM} at $x_M = 0.1$, and of G_{WW} at $x_M = 0.9$, suffer from large uncertainties. This indicates that the KBIs are not fully converged and much longer sampling is needed.[22] For the purpose of this study we compare the FG model with the different CG models within the range of these uncertainties, being aware that a quantitative match might be difficult to achieve.

4.2 Effective potentials for the different implicit solvent models

The implicit solvent models are generated as described in the simulation details. The final CG potentials are presented in figure 3. Figure 3 a) shows the final effective pair potentials for the methanol-methanol (MM) interaction. As theoretically derived in the section on the equality between IMC and REO, REO (dashed green line) and IMC (solid red line) lead to very nearly the same set of pair potentials. The REO LD model shows a slightly smaller second maximum in the pair potential, which is illustrated through the dashed blue curve

Table 2: Kirkwood-Buff Integrals for methanol-methanol (MM), methanol-water(MW) and water-water (WW) at different methanol mole fractions (x_M) obtained by averaging RKBIs and GKBIs between 1.0 and 1.5 nm.

G_{ij}	x_M	RKBI (nm^3)	GKBI (nm^3)
G_{WW}	0.1	-0.026 $\pm$ 2.2$\cdot 10^{-3}$	-0.025 $\pm$ 2.5$\cdot 10^{-3}$
	0.5	0.021 $\pm$ 3.0$\cdot 10^{-3}$	0.025 $\pm$ 2.5 $\cdot 10^{-3}$
	0.9	-0.02 $\pm$ 1.97$\cdot 10^{-2}$	0.01 $\pm$ 1.05 $\cdot 10^{-2}$
G_{MW}	0.1	-0.039 $\pm$ 8.1 $\cdot 10^{-3}$	-0.043 $\pm$ 8.7 $\cdot 10^{-3}$
	0.5	-0.041 $\pm$ 1.3$\cdot 10^{-3}$	-0.043 $\pm$ 1.0 $\cdot 10^{-3}$
	0.9	-0.016 $\pm$ 0.3 $\cdot 10^{-3}$	-0.016 $\pm$ 0.4 $\cdot 10^{-3}$
G_{MM}	0.1	-0.16 $\pm$ 3.58 $\cdot 10^{-2}$	-0.14 $\pm$ 3.70 $\cdot 10^{-2}$
	0.5	-0.078 $\pm$ 0.3$\cdot 10^{-3}$	-0.077 $\pm$ 0.4 $\cdot 10^{-3}$
	0.9	-0.075 $\pm$ 0.2 $\cdot 10^{-3}$	-0.075 $\pm$ 0.2 $\cdot 10^{-3}$

in figure 3 a). This shift is caused by the additional MM LD potential, presented in figure 3 c). This potential is small in magnitude, but weakly attractive and lowers the potential barrier in the pair potential. Notably, the small magnitude of the LD potential reveals only a weak multibody nature of the MM interaction at this state point.

The CG potentials for water in implicit methanol are shown in figure 3 b). Similarly to CG methanol in implicit water, the effective pair potentials obtained from IMC (solid red line) and from the pair only relative entropy optimization (dashed green line) nearly overlap, with a small difference between the first maximum and the second minimum. Interestingly, the pair potential of the REO LD model (dashed blue line) lacks the inner potential well present in the other two models, but the absence of an attractive well is compensated by the additional WW LD potential (figure 3 d)). This points towards a strong coupling between the pair and LD potentials, which may mean that the attractive interactions are more naturally captured at a mean-field multibody level (LD potential), but this result may also be an outcome of an overlap in the function space of the pair and LD potentials whereby either can compensate for the another in a manner to which the CG optimization procedure is insensitive. Indeed similar compensation effects were reported by Sanyal and Shell[34] and more recently by Scherer and Andrienko for the coupling between two- and three-body

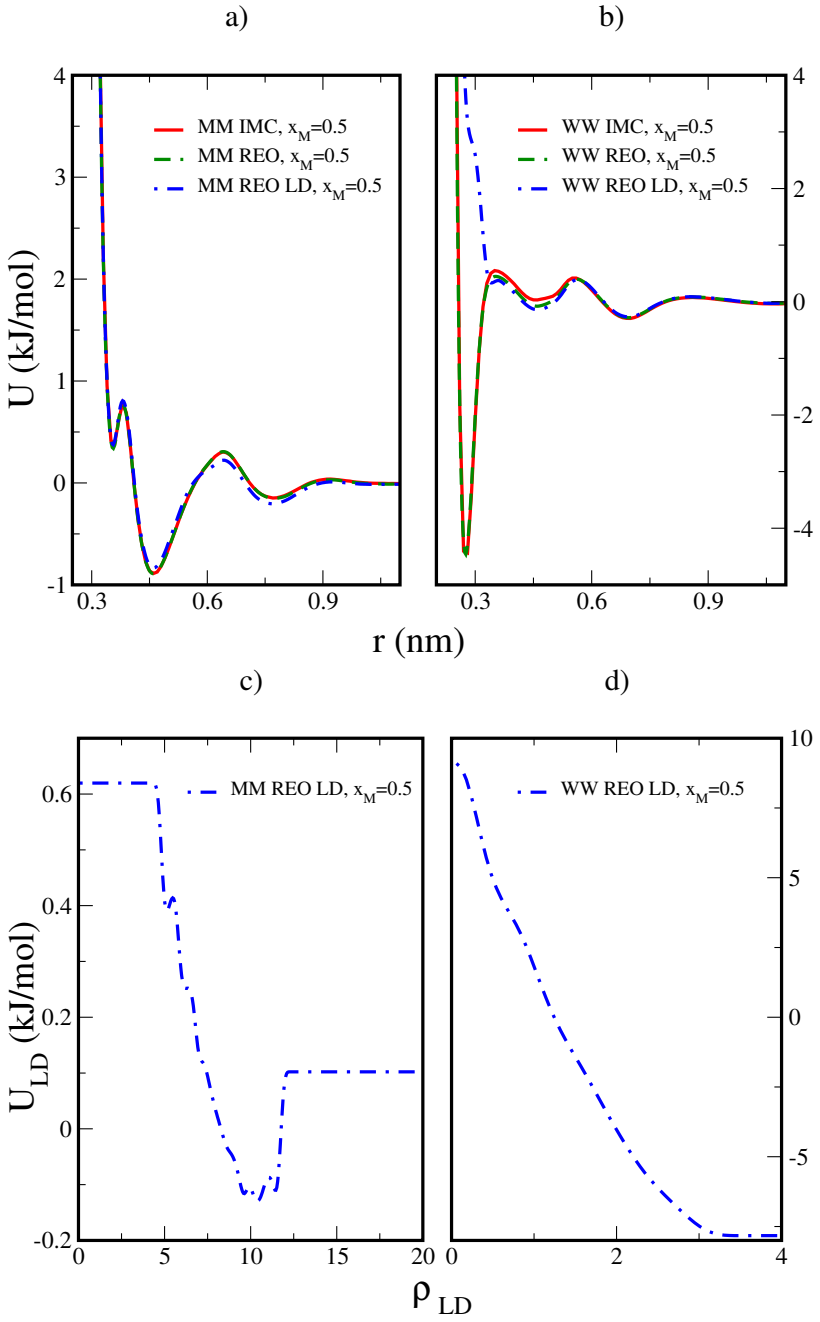


Figure 3: a) effective pair potentials for the methanol-methanol (MM) interaction at 50% methanol and b) for the water-water (WW) interaction obtained from IMC (solid red lines), REO (dashed green line) and REO LD optimization (dotted blue line); c) local density potentials for the MM interaction derived at 50% methanol and d) for the WW interaction.

interactions.[70] The WW LD potential is larger in magnitude compared to the MM one and saturates at a minimum of four neighboring water molecules, which no doubt corresponds to the preference for tetrahedral coordination in liquid water. The LD potential reaches a plateau from 4 neighbors onwards, which is by design of the relative entropy optimization. At some point the system no longer explores local densities beyond a maximal value (see for example figure 4 c)) where the largest is $\approx$ 15. So, beyond these values there is no information in the reference simulation that can be used to tune the potential. The relative entropy algorithm thus extrapolates a constant value thereafter. Noticeably, the WW IMC potentials needed 10 iterations more to converge than the MM ones (15 vs. 5 iterations).

4.3 Structural and Thermodynamic Representability

To evaluate the representability of the derived CG models, we compute the RDF between CG water in implicit methanol, and between CG methanol in implicit water, both at a mole fraction of x_M=0.5. In figure 4 a), the RDFs between methanol molecules for the different CG models in comparison with the center of mass RDF calculated from FG simulations are presented. As one sees, all CG models overlap with the FG RDF, showing two distinct maxima, one at $\approx$ 0.3 nm that stems from the methyl-hydroxyl interaction and one at $\approx$ 0.5 nm that stems from the methyl-methyl interaction. In figure 4 c), the LD distributions obtained from the REO and REO LD model are shown in comparison with the FG reference. As one sees, the REO model (dashed green line) already captures the FG LD distribution (solid black line) quite accurately and the additional LD potential improves the match only slightly (dashed blue line). This corresponds to the similarity in the underlying pair potentials and the small contribution of the LD potentials, and further indicates only a weak multibody nature of the MM interaction.

The water-water (WW) RDFs presented in figure 4 b), show a very sharp first peak that is present in all models. This points towards a high probability to find a water molecule next to another one, but due to the narrow width, only $\approx$ 4 water molecules are found in

the first solvation shell, in comparison to icosahedral coordination in simple liquids. This is in agreement with prior experimental and theoretical studies.[44,45,71] Despite the similarity in the qualitative appearance of the RDF, the REO LD model (dashed blue line) slightly underestimates this first peak, as shown in figure 4 b). The REO model (dashed green line) also slightly underestimates the first peak and the IMC model (solid red line) - by construction - shows the best agreement. This is surprising, since IMC and REO show only minor differences in the underlying pair potentials. One possible reason could be the different interpolation and extrapolation schemes applied by the two coarse graining software packages during the generation and/or evaluation of the final tabulated potentials. This assumption is supported by figure 5 a), where we compare the raw output of VOTCA, i.e. no interpolation between data points, and the two tabulated potentials used in the simulation. Here, potentials are interpolated. Moreover, the relative entropy approach utilizes piecewise cubic splines to represent interactions that are distinct from the potential interpolation in the IMC approach that uses the Akima interpolation scheme. Another reason for the differences between IMC and REO models may be the distinct convergence strategies, involving iteration in the former and minimization in the latter.

One sees that the repulsive part as well as the potential well in the tabulated IMC potential (solid black line) is slightly softer than for the REO tabulated potential (red line) and the raw output of VOTCA (green line). On the basis that structure is mainly determined by the short range part of a potential and given the narrow width of the first peak in the RDF, this small difference between raw output data and final tabulated potentials could cause the structural difference observed. For methanol-methanol, those small differences in the repulsive region are not observed as depicted in figure 5 b).

The methanol-methanol LD distributions seem insensitive to the parametrization method, consistent with the weak LD interactions that emerge in the REO LD model. The water-water LD distribution, on the other hand, reveals that the agreement between the FG model (solid black line) and the REO model (dashed green line) is improved by the LD potential

(dashed blue line), as shown in figure 4 d). Nevertheless, the improved agreement in the LD distributions comes with a loss of accuracy in the RDF, so it seems difficult to quantitatively match both at the same time with the current set of potentials. A reason for this may be the expanded parameter space for relative entropy minimization when LD potentials are included, in which spline discretization (i.e., knot density) begins to become important. This issue may also relate to the particular balancing of attractive interactions, which manifest in the LD strategy as a compensation between repulsive pair interactions and attractive LD ones, as discussed above. Indeed, in principle relative entropy minimization with pair and LD potentials should exactly reproduce both the pair and LD correlation functions, as discussed by Chaimovich and Shell,[72] such that observed differences must be due either to algorithm convergence, statistical fluctuations, or the manner by which the potential is approximated (e.g., splines of a chosen knot density).

To quantify the mismatch in the RDF, the root mean squared (RMS) error in the g(r) calculation between the FG and the CG system is computed as,

$$RMS\ \texttt{error} = \sqrt{\sum_{i=1}^{N}(g^{FG}(r) - g_i^{CG}(r))^2} \tag{33}$$

by performing the summation over a total of $N = 5$ simulations. The results are presented in figure 6. The error for the MM interaction, which is depicted in figure 6 a), is less than 0.01 for all the CG models, which further confirms the accurate structural quality of all CG methanol models in implicit water. However, for CG water the REO LD model (see figure 6 b)) shows a deviation of about 0.4 at the short distances, which corresponds to the underestimation of the first peak in the g(r) presented in figure 4 b). The IMC and REO models are very similar with respect to the RMS error and perform much better in terms of structural agreement compared to the REO LD model.

To further assess the quality of the derived implicit solvent CG models, the average GKBI values are computed for all models and compared to the FG system (see table 3).

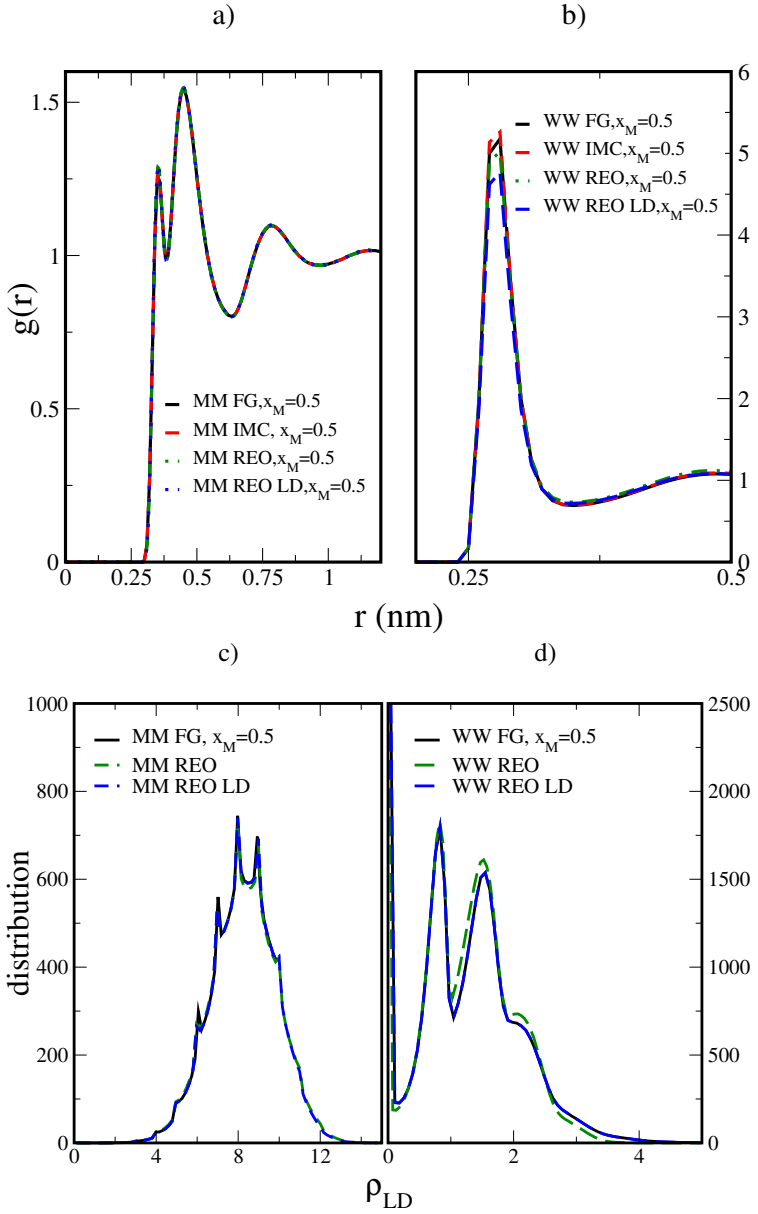


Figure 4: Representability analysis of the RDFs for the implicit solvent models at $x_M = 0.5$: a) Comparison of center of mass RDFs between methanol molecules (MM); b) Comparison of center of mass RDFs between water molecules (WW); Representability analysis of the LD distributions for the REO implicit solvent models at $x_M = 0.5$: c) Comparison of the LD distributions of methanol molecules (MM); d) Comparison of the LD distributions of water molecules (WW). The FG model is illustrated through the solid black line, the IMC model through the dashed red line, the REO model through the dotted green line and the REO LD model through the dotted blue line.

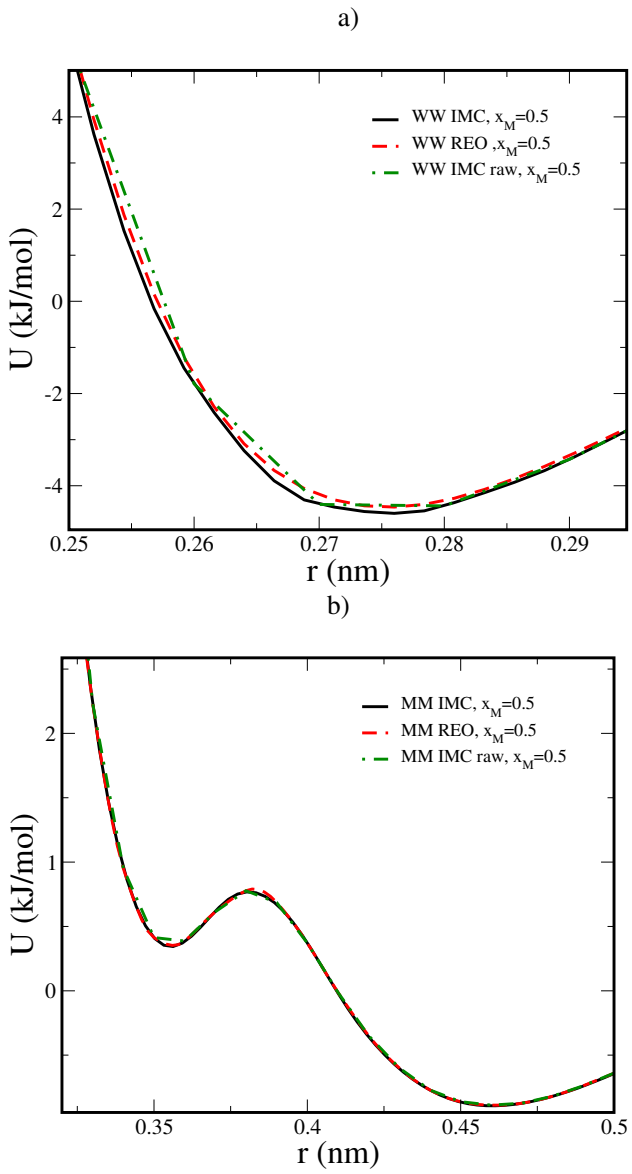


Figure 5: Comparison of final tabulated pair potentials between IMC (black line), REO (red line) and the raw output of the VOTCA simulation package (green line). a) for the water-water interaction; b) for the methanol-methanol interaction.

Note that these assessments are intimately connected to a model's representation of the pair correlation functions, which all three methods should reproduce exactly in the limit of an infinite number of pair potential bins or spline points, according to the considerations in the section on the equality between IMC and REO. As such, differences in GKBI values between models necessarily represent differences in CG algorithm convergence properties and/or particulars of numerical representation of the potentials.

The GKBIs of all CG methanol models, G_{MM}, are in good agreement with the underlying FG one. This is in agreement with the structural overlap between all models. All CG water models show larger GKBI values compared to the FG reference. This general increase in the GKBIs for CG water in implicit methanol implies an effective stronger affinity between water molecules outside the first solvation shell. As a consequence, all CG water models are more compressible. The difference between the IMC and REO and REO LD model coincides with the differences observed in the RDFs. The REO and REO LD model show the same value within error bars. In general, inclusion of the LD potentials does not seem to impact the GKBI values, with respect to relative entropy pair-only models. This is likely due to the emphasis on long range correlations of the GKBI integrals and the fact that the LD potential modifies short range interactions.

Table 3: Ganguly corrected Running Kirkwood-Buff Integrals averaged between 1.0 and 1.5 nm for the different CG models

X_M	model	G_{WW} (nm^3)	G_{MM} (nm^3)
0.5	FG	$0.025 \pm 2.5 \cdot 10^{-3}$	$-0.077 \pm 0.4 \cdot 10^{-3}$
	IMC	$0.094 \pm 0.9 \cdot 10^{-3}$	$-0.078 \pm 0.2 \cdot 10^{-3}$
	REO	$0.063 \pm 2.2 \cdot 10^{-3}$	$-0.076 \pm 0.2 \cdot 10^{-3}$
	REO LD	$0.065 \pm 1.1 \cdot 10^{-3}$	$-0.074 \pm 0.1 \cdot 10^{-3}$

The similarity in the GKBIs between the REO and REO LD models might further point towards a dominant role of direct pairwise interactions in the system. To probe this idea, we compute the cluster size distribution that is an emergent property dependent on the collective interactions in the system. If multibody effects are significant, then one expects the

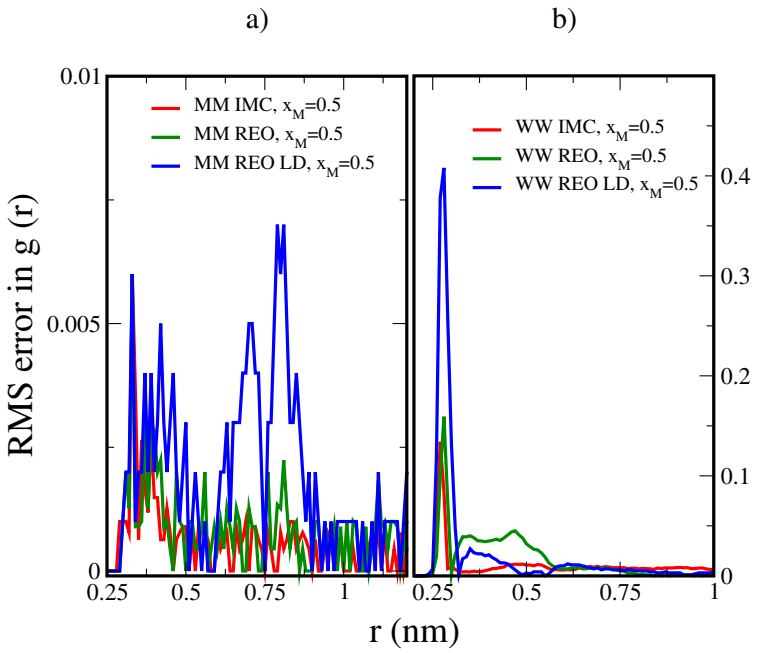


Figure 6: RMS error between FG and CG RDFs for the implicit solvent models at $x_M = 0.5$: a) RMS error in the methanol methanol (MM) RDF; b) RMS error in the water water (WW) RDF. The difference between the FG and the IMC model is illustrated through the solid red line, between the FG and the REO model through the solid green line and between the FG and the REO LD model through the solid blue line.

cluster size distribution to be impacted by the ability of the CG model to reproduce them.[32] To evaluate the cluster size distributions, we chose the same cut-off values as for the LD potentials to determine whether or not particles are in contact and thus belong to the same cluster. In practice, the cluster size distribution can be very sensitive to the cutoff choice. As one sees in figure 7 a), the FG methanol molecules comprise a single, system-spanning cluster at $x_M = 0.5$. A corresponding snapshot of the FG system presented in figure 7 a) confirms this picture. Note that for visual clarity we do not show the water molecules, which are explicitly present in the simulation. The network-like structure is similar to the formation of rings and strings, which follows the observations made by Guo et al. and Perera et al..[41,45] Since all CG models are able to match the cluster size distribution of the underlying FG system, it suggests that multibody effects play a less significant role in forming this percolating structure.

Figure 7 b) shows the cluster size distribution of water molecules for the FG model and the implicit methanol models. Here, no large water clusters are formed and the water molecules largely populate isolated clusters as well as doublets and triplets, as indicated from the snapshot presented in the inset (methanol molecules are not depicted in the snapshot for visual clarity). The network-like methanol structure restricts the number of water-water contacts. At this concentration, water molecules are not significantly tetrahedrally coordinated, which supports the picture that the LD potential has a less attractive impact on the performance of the model compared to the REO or IMC model, since the number of neighboring molecules is most likely to be < 3.

Further, the formation of only small aggregates points towards a less dominant role of multibody effects, at least at the level of water-water interactions, and implies a weakened effect of the LD potentials. That observation most likely explains why the REO LD model is slightly less populated at low coordination numbers and cluster sizes. This observations coincides with the work of Laaksonen et al., who showed a loss in tetrahedral coordination of water molecules from $x_M = 0.5$ onwards.[44] The difference between the REO and REO LD model

can be explained by an increased attraction between water molecules introduced by the negative tail of the LD potential if the number of neighboring waters is ≥ 3. This tail has a lower value than the minimum in the REO model and thus introduces stronger attraction.

4.4 Structural and Thermodynamic transferability

Despite the negligible contribution of LD potentials to the representability of the CG models at $x_M = 0.5$, we now test their effect on the transferability towards different methanol mole fractions. Because the IMC and REO optimized pair models are nearly identical, we focus on comparison of the IMC and the REO LD models to study the effect of LD potentials on the transferability of the derived CG models.

In the top panel of figure 8, the RDFs for the two CG models are shown at a mole fraction of $x_M = 0.1$ in comparison with the FG RDF at $x_M = 0.1$ and at $x_M = 0.5$. In figure 8 a), one can see that at $x_M = 0.1$ the CG models consistently show a decrease in the first peak of the MM RDF relative to the FG one at the same composition, and lower than the FG structure at $x_M = 0.5$. Further, the CG models underestimate the second maximum and overestimate the second minimum relative to the FG simulation at the same composition. Despite this quantitative mismatch, both CG models reproduce the main features (maxima, minima) of the FG structure well and show some aspects of transferability, when compared to the reference structure at $x_M = 0.5$.

For water, the REO LD model (solid light blue line) closely matches the structure of the FG system at high water content (solid black line), as depicted in figure 8 b). In contrast, the IMC water model fails to capture the g(r) structure and instead better seems to match the FG structure at $x_M = 0.5$ (solid orange line) that was the original parametrization condition. This may point to the role of multibody water-water interactions as relevant to the solution behavior at high water concentrations.

The bottom panel of figure 8 shows the RDFs at $x_M = 0.9$. In figure 8 c), the CG model for methanol in implicit water shows no significant difference between the IMC model (solid red

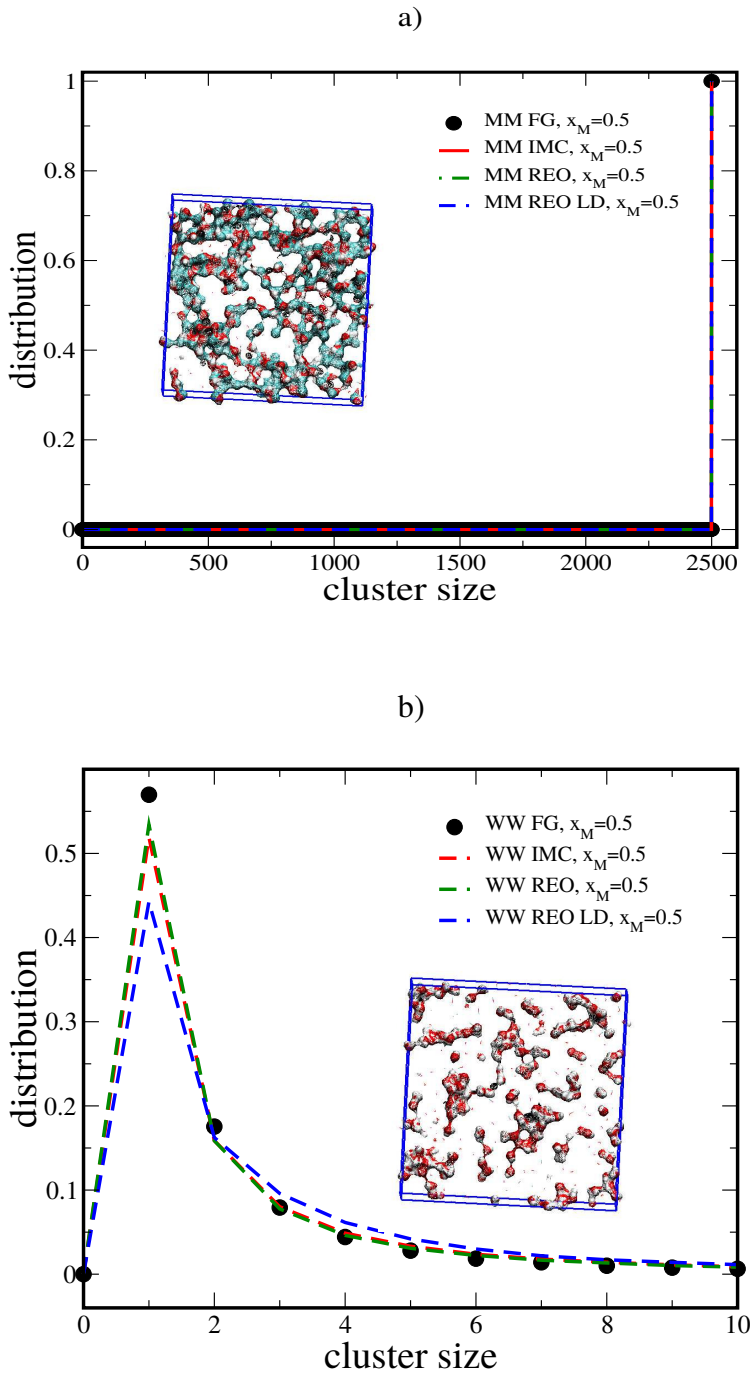


Figure 7: Representability analysis of the cluster size distributions for the implicit solvent models at $x_M = 0.5$: a) Comparison between the cluster size distribution of methanol molecules (MM). The inset depicts a snapshot of the FG trajectory without the water molecules for visual clarity; b) Comparison between the cluster size distribution of water molecules (WW). The FG model is illustrated through the black circles, the IMC model through the dashed red line, the REO model through the dotted green line and the REO LD model through the dashed blue line. The inset depicts a snapshot of the FG trajectory without the methanol molecules for visual clarity.

line) and the REO LD model (dotted light blue line). Similar to the lower mole fraction case, both of these models cannot reproduce exactly the FG structure (solid black line), but are able to capture the location of the first maximum in the RDF compared to the structure at at $x_M = 0.5$ (solid orange line). In figure 8 d), one sees that the IMC model (solid red line) leads to a decreased first peak in the RDF for CG water in implicit methanol compared to the FG model (solid black line). Further, the RDF obtained from the IMC model overlaps with the RDF of the FG system (dashed orange line) at the reference point ($x_M = 0.5$). These results suggest that IMC (and relative entropy minimization with pair potentials only) embed composition-specific interactions and correlations that propagate to other compositions. They also suggest that local density interactions can improve transferability when multibody interactions are present. On the other hand, the presence of the LD potential (solid light blue line) leads to a stronger water-water aggregation and a significantly enhanced first peak. It is likely in this case that multibody water interactions are over emphasized in dilute solutions where each water has few neighbors and the LD potential at low water-water coordination is inappropriate, having been constructed from a more water-rich reference.

In general, it seems that to model the structural transferability, LD dependent potentials play no significant role for methanol in implicit water. Contrarily, for water in implicit methanol, the presence of a LD potential improves the structural transferability with decreasing methanol concentration. This is in agreement with the observations of Sanyal and Shell,[34] who pointed out that water-water LD potentials lead to transferable CG models of aqueous mixtures only when water-water interactions have a major contribution to the multibody PMF. Further, this corresponds to the work of Laaksonen et al.,[44] who found out that the dominant species in the system is mainly responsible for the structural features of the mixture. Following this argument, the decreasing effect of LD contributions with increasing methanol concentration reveals the difficulty of using LD potentials to effectively capture the unusual mixing behavior of methanol and water upon increasing methanol mole fraction.

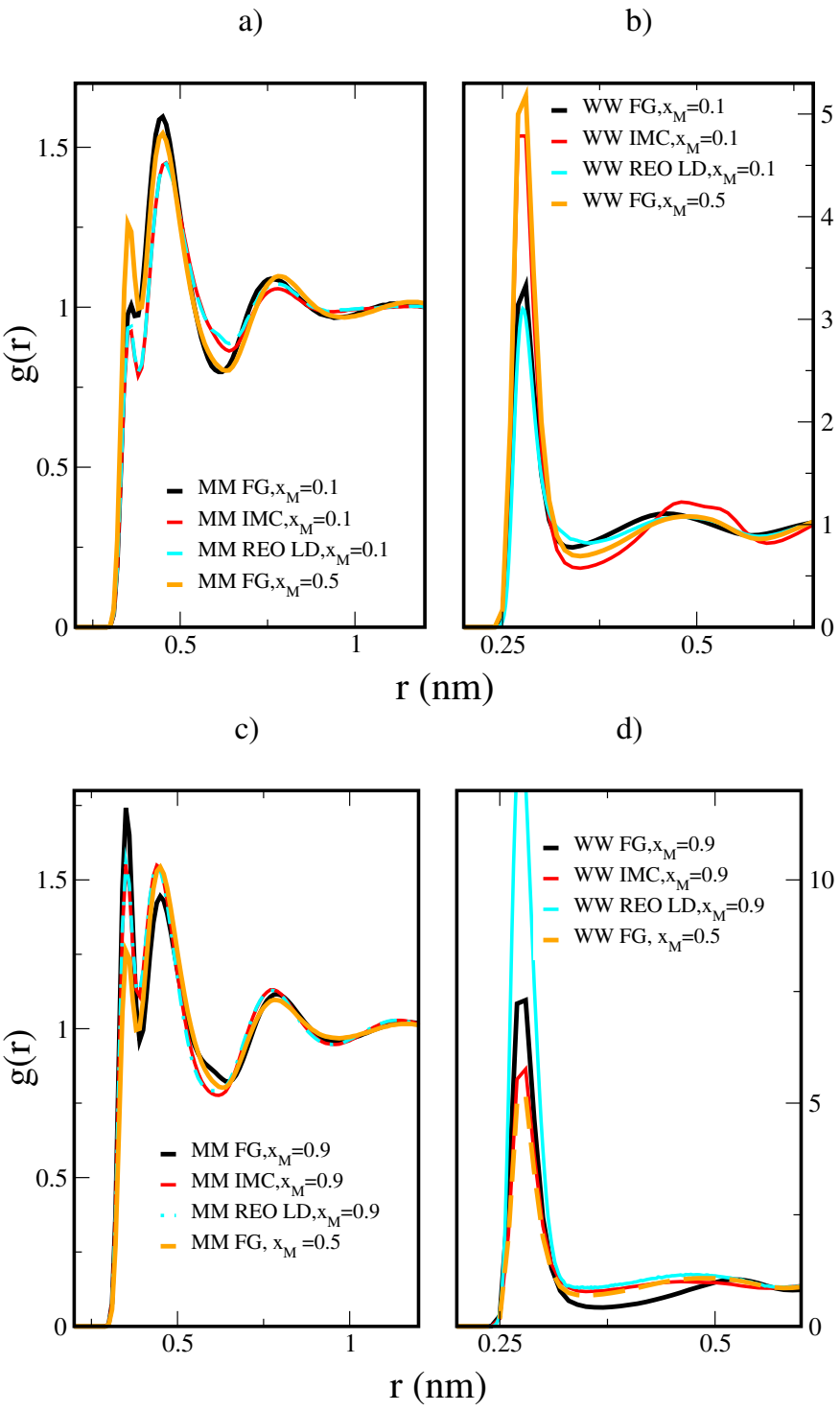
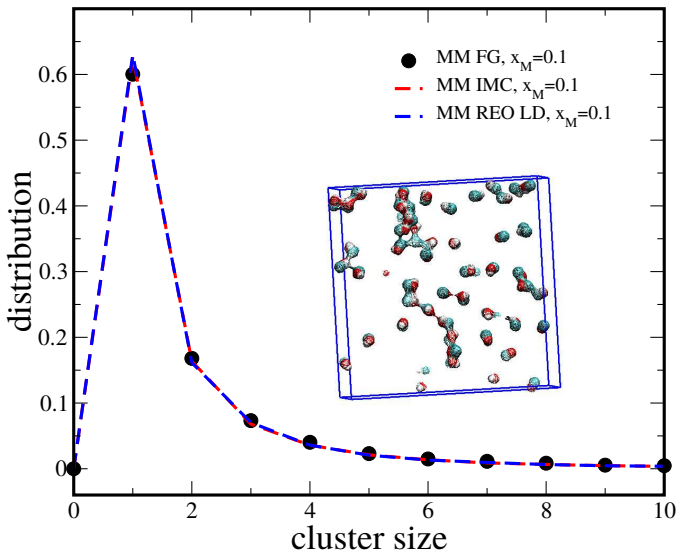


Figure 8: a) Comparison of the center of mass RDFs at $x_M = 0.1$ between methanol molecules (MM); b) Comparison of the center of mass RDFs at $x_M = 0.1$ between water molecules (WW); c) Comparison of the center of mass RDFs at $x_M = 0.9$ between methanol molecules (MM); d) Comparison of the center of mass RDFs at $x_M = 0.9$ between water molecules (WW). The FG model is illustrated through a solid black line, the IMC model is illustrated through the solid red line and the REO LD model is illustrated through the solid/dotted light blue line and the FG reference at $x_M = 0.5$ is illustrated through the solid orange line.

The negligible effect of multibody contributions is again illustrated by the methanol cluster size distribution for non-reference cases, as presented in figure 9. In figure 9 a), one sees that at low methanol concentrations only a few methanol molecules are grouped together, whereas at high concentrations (figure 9 b)) there exists only a single cluster composed of all methanol molecules in the system, which is qualitatively similar to x_M=0.5. These observations are further supported by the snapshots of the FG system depicted in the inset of the figures, where water molecules are not shown for visual clarity. There again a network-like structure is visible for methanol mole fractions > 0.5. Since all CG models perform in the same way, it further indicates that methanol-methanol interactions are not dominated by multibody effects in aqueous solutions.

We now turn to the cluster size distributions in the implicit-methanol systems. In figure 10 a), the distribution for waters at $x_M = 0.1$ shows two distinct cluster sizes for the FG system. On the one hand, a cluster emerges with very few water molecules, while on the other hand one appears with the number of water molecules in the system (i.e., a percolating or system spanning cluster). A snapshot of the FG system (shown in the inset) illustrates this, where the methanol molecules are not depicted for visual clarity. One sees a smaller cluster in the lower left corner and a large, system spanning cluster. It appears that small methanol sub-domains, present at low methanol concentration ($x_M = 0.1$, see figure 9 a)), limit that all of the water aggregates into a single cluster. This produces the separation of peaks in the WW cluster distribution (see figure 10 a)) and is in agreement with the study of Perera et al..[45] The IMC model produces clusters with only few water molecules, that is more similar to the cluster distributions at $x_M = 0.5$. This is in agreement to what we observe for the g(r) as well, where the g(r) of the IMC model is more similar to the one of the FG model at $x_M = 0.5$ than at $x_M = 0.1$. Thus, this model is not transferable towards lower methanol concentrations. Interestingly, the REO LD model cluster distribution attains better agreement with the cluster size distribution, in particular capturing the two-state population at low and high cluster size. In this case, we adjusted the distance cut-off to

a)



b)

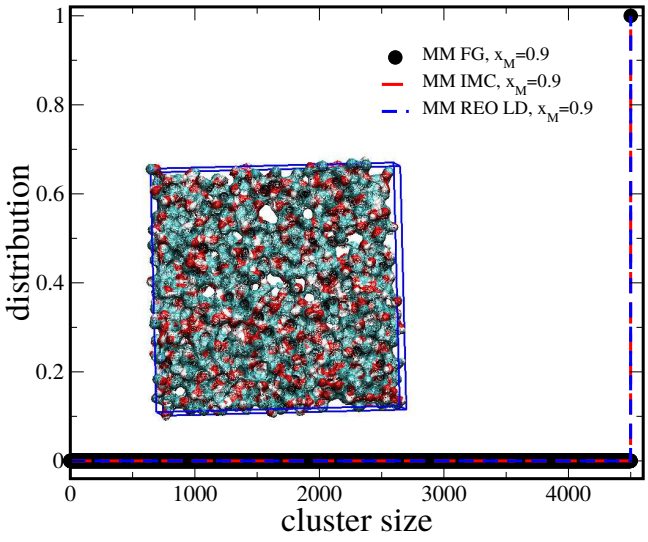


Figure 9: Transferability analysis of the cluster size distributions for the implicit solvent models: a): Comparison of cluster size distribution of methanol molecules (MM) at $x_M = 0.1$; b) Comparison at $x_M = 0.9$. The FG model is illustrated through black circles, the IMC model through the dashed red line and the REO LD model is illustrated through the dashed blue line. The insets depict a snapshot of the corresponding FG trajectories without the water molecules for visual clarity.

0.37 nm to effectively locate water neighbors in this heterogeneous environment. The cluster size distribution at $x_M = 0.1$ clearly shows that the LD potential improves the CG water model. Moreover, it confirms the picture that in water rich phases, multibody effects play a significant role and LD dependent potentials are effective strategies for capturing such interactions. At high methanol concentrations ($x_M = 0.9$), no large clusters of water are formed and the LD dependent potential does not improve the CG model compared to the IMC one. The more isolated behavior of water molecules is again depicted in a snapshot of the corresponding molecules taken from the FG system. This indicates a negligible role of water and multibody effects in methanol rich phases, which corresponds to the picture of Laaksonen and Perera.[44,45]

Another prediction of LD potentials is further indicated by the GKBIs presented in table 4. At low concentrations, both the IMC and REO LD model, show stronger affinity between methanol molecules than the FG model does. The LD potential even emphasizes this effect. The same trend can be observed at $x_M = 0.9$. Here, the REO LD model shows again the largest GKBI value, whereas the IMC model is closer to the FG system. This suggests further that LD effects do not play a significant role on the behavior of methanol in aqueous solutions.

Contrary to that, the REO LD model reproduces the FG WW-GKBI at low methanol concentrations ($x_M = 0.1$). At $x_M = 0.9$, the WW-GKBI is however overestimated by the REO LD model, as indicated by the RDF as well. Thus, the GKBI analysis follows the same trends observed in the RDFs. This brings us to the following conclusion: in water rich phases, where LD effects have a larger contribution to the water-water multibody PMF, LD potentials improve the transferability of the derived CG models. However, with increasing methanol concentrations, these contributions become less significant and the CG models do not further improve.

However, it is essential to note that the performance of all models in reproducing GKBIs is purely algorithmic in nature because the IMC, REO, and REO LD methods all formally

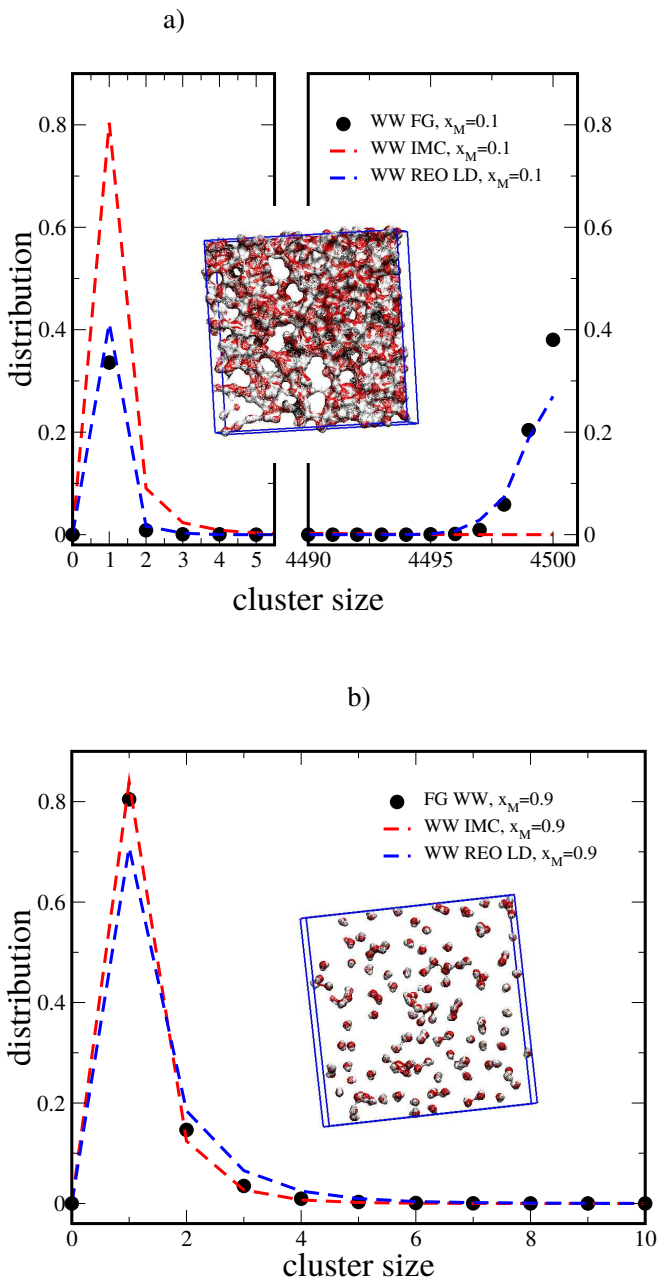


Figure 10: Transferability analysis of the cluster size distributions for the implicit solvent models: a) Comparison between the cluster size distribution of water molecules (WW) at x_M=0.1; b) Comparison between the cluster size distribution of water molecules (WW) at x_M=0.9. The FG model is illustrated through black circles, the IMC model through the dashed red line and the REO LD model is illustrated through the dashed blue line. The insets depict a snapshot of the corresponding FG trajectories without the methanol molecules for visual clarity.

should (in principle) locate CG models that reproduce exact g(r) forms upon which the GKBIs rely. Therefore, any difference between the properties in table 4 must be considered a result of numerical approximations (e.g., tabulated or splined potentials) or convergence (CG method), and the fact that the GKBI integrals are sensitive to subtle pair correlation effects at large distances.

Table 4: Average Ganguly corrected Running Kirkwood-Buff Integrals at different concentrations of methanol obtained either from FG simulations or from CG simulations with previously generated potentials at $x_M = 0.5$.

x_M	model	G_{WW} (nm^3)	G_{MM} (nm^3)
0.1	FG	-0.025 $\pm$ 2.5$\cdot 10^{-3}$	-0.14 $\pm$ 3.70 $\cdot 10^{-2}$
	IMC	0.090 $\pm$ 1.4$\cdot 10^{-3}$	-0.091 $\pm$ 1.2$\cdot 10^{-3}$
	REO LD	-0.026 $\pm$ 0.2$\cdot 10^{-3}$	-0.054 $\pm$ 0.2$\cdot 10^{-3}$
0.9	FG	0.01 $\pm$ 1.05$\cdot 10^{-2}$	-0.075 $\pm$ 0.3$\cdot 10^{-3}$
	IMC	0.084 $\pm$ 3.5$\cdot 10^{-3}$	-0.071 $\pm$ 0.2$\cdot 10^{-3}$
	REO LD	0.380 $\pm$ 0.6$\cdot 10^{-3}$	-0.066 $\pm$ 0.2$\cdot 10^{-3}$

5 Conclusion

In this work, we addressed the question of whether single site CG models for methanol and water mixtures, in an implicit solvent environment, can capture the mixing behavior of these two components. We could have also considered mapping methanol to a two-bead model to account for non-polar and polar sites. But, this would require up to 6 local density potentials for the combination of central and neighbor type. This would make the model (and its parameterization) significantly more complex, which we wanted to avoid. We derived implicit solvent CG models for various water/methanol mixtures by application of two different coarse graining methods, namely IMC and relative entropy optimization. We showed analytically and numerically, that under the assumptions of a discretized Hamiltonian and the application of a Newton-Raphson scheme, the methods are formally equal.

We further investigated if embedding of LD potentials improves the derived CG models

in terms of concentration transferability. In agreement with previous work by Sanyal and Shell[32,34] we found, if water-water LD contributions play a dominant role in the multibody PMF, LD potentials improve the transferability of water in implicit methanol towards lower methanol concentrations. Unlike the work in [32], here we find that LD potentials do not improve transferability in water-lean solutions. The LD included CG model for methanol in implicit water shows no significant improvement at either concentrations. When compared to a CG model with pair potentials only, the presence of a LD potential even emphasizes the affinity between methanol molecules as indicated by the GKBI. One difference in the present study from [34], which showed that LD models were consistently able to improve transferability across composition space in benzene-water mixtures, is that here a species is made implicit. It may be that LD potentials are more effective when all species are explicitly present but coarse-grained. The interpretation that multibody effects have a less significant effect on the structural arrangement of methanol molecules is further supported by the analysis of the cluster size distributions in the systems. Here, all CG models show similar size distributions for all concentrations independent of the presence of a LD dependent potential. The CG methanol models show more network-like structure due to large volume occupancy in the system, whereas the CG water stays mostly isolated, perhaps due to restraints posed by the methanol network. In agreement with our work, Scherer and Andrienko recently found a negligible effect of multibody contributions on methanol-methanol interactions by investigating the impact of three-body potentials on the pair structure of liquid methanol.[70] What seems to be more important than multibody contributions is the ability to form hydrogen bond networks in order to accurately describe water/methanol mixtures as discussed in the literature.[41,71] Explicit electrostatics or any directionality introduced to the model, as for example done on the basis of point multipole and Gay-Berne potentials,[73,74] seems to be necessary, to accomplish the effect of hydrogen bonding. Without performing extensive analysis on the FG system, we used bottom-up coarse graining techniques to identify the negligible contribution of LD dependent potentials on structural changes in alcohol/water

mixtures. This does not mean that LD potentials cannot improve the transferability of CG models as nicely demonstrated in the literature.[32–34] It rather shows that one should carefully consider the underlying physics of the system, specifically the extent of multibody effects, for a successful application of LD potentials. Further, this study highlights the aid of bottom-up coarse graining to identify important degrees of freedom, e.g. here the ability to explicitly form hydrogen bonds, to accurately describe a system in the CG configuration space.

Acknowledgement

NvdV and DR acknowledge funding by the German Research Foundation (DFG) within the SFB-TRR 146 grant on "Multiscale Simulation Methods for Soft Matter Systems". TS and MSS acknowledge funding from the National Science Foundation through grant no. CHEM-1300770

References

(1) 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.

(2) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.

(3) 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*, 2108–2119.

(4) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *EPL* **1994**, *26*, 583–588.

(5) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.

(6) 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.

(7) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *J. Chem. Phys.* **2008**, *128*, 244115.

(8) Mullinax, J. W.; Noid, W. G. Generalized Yvon-Born-Green Theory for Molecular Systems. *Phys. Rev. Lett.* **2009**, *103*.

(9) 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.

(10) Schommers, W. Pair potentials in disordered many-particle systems: A study for liquid gallium. *Phys. Rev. A* **1983**, *28*, 3599–3605.

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

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

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

(14) Louis, A. A. Beware of density dependent pair potentials. *J. Phys.: Condens. Matter* **2002**, *14*, 9187–9206.

(15) Johnson, M. E.; Head-Gordon, T.; Louis, A. A. Representability problems for coarse-grained water potentials. *J. Chem. Phys.* **2007**, *126*, 144509.

(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.

(17) Guenza, M. Thermodynamic consistency and other challenges in coarse-graining models. *Eur. Phys. J. ST* **2015**, *224*, 2177–2191.

(18) Wang, H.; Junghans, C.; Kremer, K. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *Eur. Phys. J. E* **2009**, *28*, 221–229.

(19) Ghosh, J.; Faller, R. State point dependence of systematically coarse-grained potentials. *Mol. Simul.* **2007**, *33*, 759–767.

(20) Allen, E. C.; Rutledge, G. C. A novel algorithm for creating coarse-grained, density dependent implicit solvent models. *J. Chem. Phys.* **2008**, *128*, 154115.

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

(22) Ganguly, P.; van der Vegt, N. F. A. Convergence of Sampling Kirkwood-Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 1347–1355.

(23) Rudzinski, J. F.; Noid, W. G. Coarse-graining entropy, forces, and structures. *J. Chem. Phys.* **2011**, *135*, 214101.

(24) Wagner, J. W.; Dama, J. F.; Durumeric, A. E. P.; Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys.* **2016**, *145*, 044108.

(25) 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.

(26) Murtola, T.; Falck, E.; Karttunen, M.; Vattulainen, I. Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints. *J. Chem. Phys.* **2007**, *126*, 075101.

(27) Moradzadeh, A.; Motevaselian, M. H.; Mashayak, S. Y.; Aluru, N. R. Coarse-Grained Force Field for Imidazolium-Based Ionic Liquids. *J. Chem. Theory Comput.* **2018**, *14*, 3252–3261.

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

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

(30) Das, A.; Andersen, H. C. The multiscale coarse-graining method. V. Isothermal-isobaric ensemble. *J. Chem. Phys.* **2010**, *132*, 164106.

(31) 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.

(32) Sanyal, T.; Shell, M. S. Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation. *J. Chem. Phys.* **2016**, *145*, 034109.

(33) DeLyser, M. R.; Noid, W. G. Extending pressure-matching to inhomogeneous systems via local-density potentials. *J. Chem. Phys.* **2017**, *147*, 134111.

(34) Sanyal, T.; Shell, M. S. Transferable Coarse-Grained Models of Liquid-Liquid Equilibrium Using Local Density Potentials Optimized with the Relative Entropy. *J. Phys. Chem. B* **2018**, *122*, 5678–5693.

(35) Rosenberger, D.; van der Vegt, N. F. A. Addressing the temperature transferability of structure based coarse graining models. *Phys. Chem. Chem. Phys.* **2018**, *20*, 6617–6628.

(36) Allen, E. C.; Rutledge, G. C. Evaluating the transferability of coarse-grained, density-dependent implicit solvent models to mixtures and chains. *J. Chem. Phys.* **2009**, *130*, 034904.

(37) Izvekov, S.; Chung, P. W.; Rice, B. M. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. *J. Chem. Phys.* **2010**, *133*, 064109.

(38) Tanaka, H.; Gubbins, K. E. Structure and thermodynamic properties of water-methanol mixtures: Role of the water-water interaction. *J. Chem. Phys.* **1992**, *97*, 2626–2634.

(39) Soper, A. K.; Finney, J. L. Hydration of methanol in aqueous solution. *Phys. Rev. Lett.* **1993**, *71*, 4346–4349.

(40) Scatena, L. F. Water at Hydrophobic Surfaces: Weak Hydrogen Bonding and Strong Orientation Effects. *Science* **2001**, *292*, 908–912.

(41) Guo, J.-H.; Luo, Y.; Augustsson, A.; Kashtanov, S.; Rubensson, J.-E.; Shuh, D. K.; Ågren, H.; Nordgren, J. Molecular Structure of Alcohol-Water Mixtures. *Phys. Rev. Lett.* **2003**, *91*.

(42) Soper, A. K.; Dougan, L.; Crain, J.; Finney, J. L. Excess Entropy in Alcohol-Water Solutions: A Simple Clustering Explanation . *J. Phys. Chem. B* **2006**, *110*, 3472–3476.

(43) Laage, D.; Stirnemann, G.; Hynes, J. T. Why Water Reorientation Slows without

Iceberg Formation around Hydrophobic Solutes. *J. Phys. Chem. B* **2009**, *113*, 2428–2435.

(44) Laaksonen, A.; Kusalik, P. G.; Svishchev, I. M. Three-Dimensional Structure in Water-Methanol Mixtures. *J. Phys. Chem. A* **1997**, *101*, 5910–5918.

(45) Perera, A.; Zorani, L.; Sokoli, F.; Mazighi, R. A comparative Molecular Dynamics study of water-methanol and acetone-methanol mixtures. *J. Mol. Liq.* **2011**, *159*, 52–59.

(46) Pascal, T. A.; Goddard, W. A. Hydrophobic Segregation, Phase Transitions and the Anomalous Thermodynamics of Water/Methanol Mixtures. *J. Phys. Chem. B* **2012**, *116*, 13905–13912.

(47) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. Versatile Object-Oriented Toolkit for Coarse-Graining Applications. *J. Chem. Theory Comput.* **2009**, *5*, 3211–3223.

(48) Lyubartsev, A.; Mirzoev, A.; Chen, L.; Laaksonen, A. Systematic coarse-graining of molecular models by the Newton inversion method. *Faraday Discuss.* **2010**, *144*, 43–56.

(49) Rosenberger, D.; Hanke, M.; van der Vegt, N. F. Comparison of iterative inverse coarse-graining methods. *Eur. Phys. J. ST* **2016**, *225*, 1323–1345.

(50) Carmichael, S. P.; Shell, M. S. A New Multiscale Algorithm and Its Application to Coarse-Grained Peptide Models for Self-Assembly. *J. Phys. Chem. B* **2012**, *116*, 8383–8393.

(51) Shell, M. S. In *Advances in Chemical Physics*; Rice, S. A., Dinner, A. R., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2016; pp 395–441.

(52) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(53) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(54) Mashayak, S. Y.; Jochum, M. N.; Koschke, K.; Aluru, N. R.; Rühle, V.; Junghans, C. Relative Entropy and Optimization-Driven Coarse-Graining Methods in VOTCA. *PLOS ONE* **2015**, *10*, e0131754.

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

(56) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(57) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(58) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Methanol and Aqueous Methanol Solutions. *J. Phys. Chem. B* **2005**, *109*, 15080–15086.

(59) 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.

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

(61) 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.

(62) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

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

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

(65) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions. I. *J. Chem. Phys.* **1951**, *19*, 774–777.

(66) Ben-Naim, A. *Molecular theory of solutions*; Oxford University Phys. Rev. Ess: Oxford, 2006; OCLC: 252677203.

(67) Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Vlugt, T. J. H.; Simon, J.-M. Kirkwood-Buff Integrals for Finite Volumes. *J. Phys. Chem. Letters* **2013**, *4*, 235–238.

(68) Dawass, N.; Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Simon, J. M.; Vlugt, T. J. H. Finite-size effects of Kirkwood-Buff integrals from molecular simulations. *Mol. Simul.* **2017**, 1–14.

(69) Milzetti, J.; Nayar, D.; van der Vegt, N. F. A. Convergence of Kirkwood-Buff Integrals of Ideal and Nonideal Aqueous Solutions Using Molecular Dynamics Simulations. *J. Phys. Chem. B* **2018**, *122*, 5515–5526.

(70) Scherer, C.; Andrienko, D. Understanding three-body contributions to coarse-grained force fields. *Phys. Chem. Chem. Phys.* **2018**, *20*, 22387–22394.

(71) Dixit, S.; Crain, J.; Poon, W. C. K.; Finney, J. L.; Soper, A. K. Molecular segregation observed in a concentrated alcohol-water solution. *Nature* **2002**, *416*, 829–832.

(72) Chaimovich, A.; Shell, M. S. Coarse-graining errors and numerical optimization using a relative entropy framework. *J. Chem. Phys.* **2011**, *134*, 094112.

(73) Golubkov, P. A.; Ren, P. Generalized coarse-grained model based on point multipole and Gay-Berne potentials. *J. Chem. Phys.* **2006**, *125*, 064103.

(74) Golubkov, P. A.; Wu, J. C.; Ren, P. A transferable coarse-grained model for hydrogen-bonding liquids. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2050.

Graphical TOC Entry