



HHS Public Access

Author manuscript

J Phys Chem Lett. Author manuscript; available in PMC 2019 October 08.

Published in final edited form as:

J Phys Chem Lett. 2019 August 15; 10(16): 4549–4557. doi:10.1021/acs.jpcllett.9b01228.

Understanding Missing Entropy in Coarse-Grained sSystems: Addressing Issues of Representability and Transferability

Jaehyeok Jin, Alexander J. Pak, Gregory A. Voth*

Department of Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, 5735 S. Ellis Ave., Chicago, Illinois 60637, USA

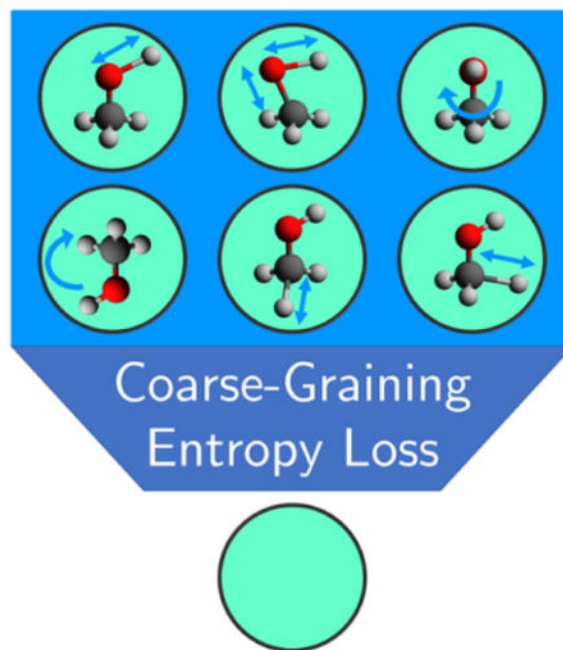
Abstract

Coarse-grained (CG) models facilitate efficient simulation of complex systems by integrating out the atomic, or fine-grained (FG), degrees of freedom. Systematically-derived CG models from FG simulations often attempt to approximate the CG potential of mean force (PMF), an inherently multidimensional and many-body quantity, using additive pairwise contributions. However, they currently lack fundamental principles that enable their extensible use across different thermodynamic state points, i.e., transferability. In this work, we investigate the explicit energy-entropy decomposition of the CG PMF as a means to construct transferable CG models. In particular, despite its high-dimensional nature, we find for liquid systems that the entropic component to the CG PMF can similarly be represented using additive pairwise contributions, which we show is highly coupled to the CG configurational entropy. This approach formally connects the missing entropy that is lost due to the CG representation, i.e., translational, rotational and vibrational modes associated with the missing degrees of freedom, to the CG entropy. By design, the present framework imparts transferable CG interactions across different temperatures due to the explicit definition of an additive entropic contribution. Furthermore, we demonstrate that transferability across composition state points, such as between bulk liquids and their mixtures, is also achieved by designing combining rules to approximate cross-interactions from bulk CG PMFs. Using the predicted CG model for liquid mixtures, structural correlations of the fitted CG model were found to corroborate a high-fidelity combining rule. Our findings elucidate the physical nature and compact representation of CG entropy and suggest a new approach for overcoming the transferability problem. We expect this approach will further extend the current view of CG modeling into predictive multiscale modeling.

Graphical Abstract:

*Corresponding author: gavoth@uchicago.edu.

Supporting Information Available: Detailed simulation settings and analysis protocols described in this work; derivation of the CG entropy expression with proof-of-concept study; convergence analysis of the suggested method; an extended theory beyond pair interactions i.e., three-body interaction; additional analysis of the theory described in this work.



Computer simulations have enabled the scientific community to understand chemical and physical phenomena at the molecular level in recent years.^{1–7} Coarse-grained (CG) models have further extended the spatiotemporal scales of simulations by removing or “integrating out” unnecessary degrees of freedom in the system of interest.^{8–15} In bottom-up CG models that are systematically parameterized from fine-grained (FG) simulations,^{15–18} the CG Hamiltonian is equivalent to the potential of mean force (PMF), and the resultant CG internal energy from summing up the CG Hamiltonian does not correspond to the FG internal energy. This is because the many-body PMF of the CG system is essentially the configuration-dependent free energy that necessarily contains an entropic contribution, i.e., it is not entirely an energetic quantity.^{19–22} However, due to its many-body nature, the configuration-dependent (e.g., pair distances) free energy depends on other thermodynamic variables (e.g., temperature, pressure, or chemical potential), and thus the so-called transferability problem limits the predictive ability of multiscale models for systems investigated in different state points.^{14, 23, 24} Despite the somewhat limited success of understanding and resolving the transferability problem using systematic approaches,^{25–32} no general method to solve the transferability problem, to our knowledge, has been proposed. Therefore, a fundamental theory to construct transferable bottom-up CG models needs to be explored.

In pursuit of this direction, a key barrier for transferability of CG interactions is the entropic contribution to the many-body PMF (free energy) of the CG variables,²⁸ as it will change throughout different state points. Since the exact form of the entropic contribution to the many-body PMF is difficult to determine, we instead endeavor to understand its relationship to the configurational entropy of the CG system. Since multiple atomic entities may be mapped into a single pseudo-atom, i.e., a CG site or “bead”, it seems clear that the entropy of the CG system is inherently reduced compared to that of the FG system since the entropic

contribution from different motions (or modes) within the CG bead are no longer resolved. This inequality is part of the general representability problem in which naïvely predicting thermodynamic quantities of the CG system based on FG expressions may result in significant deviations from the true FG values.^{18, 33–37} While such inequalities can be analytically reformulated in simple CG models,¹⁸ the generalization to physical systems with more complex interactions is quite challenging.

As such, an understanding of the CG entropy and its relation to the FG entropy is of utmost importance for not only constructing a systematic bottom-up CG model but also addressing challenges in coarse-graining, i.e., representability and transferability. We note that a previous theoretical study has addressed this issue for the first time and further derived a representability relationship with analytical analyses on global expressions for the entropy in CG systems.³⁸ A central result from Ref 38 is that the entropic contribution to the CG PMF is a conditioned relative entropy that quantifies information loss upon CG mapping. This prior work demonstrated that decreasing CG resolution increased the entropic contribution to the PMF in the case of simple models.³⁸ An important consequence of this analysis is that a practical yet accurate representation of the entropic contribution is likely necessary for transferable CG models. In this work, we aim to define and interpret the entropy within CG systems and provide a compact representation of the microscopic contributions to the CG entropy. After retrieving the missing entropy from the CG observable expression, we propose procedures for constructing transferable CG models using: (1) various temperatures and (2) different chemical environments (system transferability).

In general, the coarse-graining (CG) procedure is applied to a fine-grained (FG) system having configurations \mathbf{r}^n by applying the mapping operator \mathbf{M} : $\mathbf{r}^n \rightarrow \mathbf{R}^N$, yielding reduced configurations \mathbf{R}^N . Namely, a set of configurational states at the FG, ω_{FG} , is renormalized to the set of CG configuration states ω_{CG} . The exact configurational entropy of the CG configurational states in the canonical ensemble $S_{CG}(\omega_{CG}) = \left\langle S_{CG, \omega_{CG}}(\mathbf{R}^N) \right\rangle_{CG}$, which should be identical to the all-atom configurational entropy ($S_{AA}(\omega_{FG})$), is formulated by taking the ensemble average of the CG entropy estimator $S_{CG, \omega_{CG}}(\mathbf{R}^N)$ defined as

$$S_{CG, \omega_{CG}}(\mathbf{R}^N) = -k_B \ln p_{\omega_{CG}}(\mathbf{R}^N) - \frac{\partial W_{CG}(\mathbf{R}^N)}{\partial T} \quad (1)$$

In eq (1), $p_{\omega_{CG}}(\mathbf{R}^N)$ denotes the probability of CG configuration \mathbf{R}^N among ω_{CG} , and $w_{CG}(\mathbf{R}^n)$ is the CG many-body potential of mean force (PMF).^{21, 39} Detailed derivations for the many-body PMF case were first suggested by Ref 38 and also were given in Ref 18 – please refer to Section S1 of the Supporting Information (SI) for details. To reiterate, eq (1) shows that the naïvely observed entropy of the CG system

$S_{CG}^{\text{naïve}}(\omega_{CG}) = \left\langle -k_B \ln p_{\omega_{CG}}(\mathbf{R}^N) \right\rangle_{CG}$ is not equivalent to that of the FG entropy $S_{AA}(\omega_{FG})$.

For instance, simple polymer models can demonstrate this mismatch between FG and CG entropies where the total entropy is completely mapped to the CG model while the naïve expression yields zero entropy.¹⁸ In this sense, we denote the missing (or excess) entropy

term $\langle -W_{\text{CG}}(\mathbf{R}^N)/T \rangle_{\text{CG}}$ as the entropy (S_{map}) that is mapped into the CG system. From eq (1), estimated S_{map} by taking a canonical ensemble average (constant NVT) over CG space, such that $S_{\text{map}} = \left\langle \frac{-\partial W_{\text{CG}}(\mathbf{R}^N)}{\partial T} \right\rangle_{\text{CG}}$. In practice, it is difficult to ascertain the exact

functional form of S_{map} in complex systems, but certain approximations may prove to be useful.

Assuming that the CG PMF is adequately represented in a pairwise-decomposable form (which may not always be the case), the S_{map} term can also be decomposed in a pairwise form as below:

$$S_{\text{map}} = \left\langle \frac{-\partial W_{\text{CG}}(\mathbf{R}^N)}{\partial T} \right\rangle \approx \left\langle \sum_{IJ} \frac{-\partial \Delta W_{\text{CG}}(R_{IJ})}{\partial T} \right\rangle = \left\langle \sum_{IJ} S_{\text{map}}^{(2)}(R_{IJ}) \right\rangle \quad (2)$$

Equation (2) suggests that one can recover the missing entropy by calculating and averaging the pairwise mapping entropy $S_{\text{map}}^{(2)}(R)$, which is the temperature derivative of the CG PMF.

This is in line with the preceding work that connects the missing entropy to the mapping entropy.³⁸ In practice, we introduce a free energy decomposition scheme, which is commonly referred to as energy-entropy decomposition in eq (3) below.^{28, 40–42}

$$\Delta W_{\text{CG}}(R_{IJ}; T) = \Delta U_{\text{CG}}(R_{IJ}) - T \Delta S_{\text{CG}}(R_{IJ}) \quad (3)$$

The symbol $\Delta S_{\text{CG}}(R)$ here refers to the thermodynamic quantity in reference to R_{IJ} at infinite distance. Thus, the pairwise entropy function can be expressed as $\Delta S_{\text{CG}}(R) = -\frac{\partial \Delta W_{\text{CG}}(R; T)}{\partial T}$.

We note that we explicitly consider the energy-entropy decomposition of the many-body PMF since we restrict ourselves to the canonical ensemble in this work, such that the free energy is indeed the Helmholtz free energy. Due to the entropic contribution, the PMF has an explicit dependence on temperature, which also provides a useful framework for constructing temperature transferable CG models.^{28, 40–42} Hence, S_{map} approximately becomes $\langle \sum_{IJ} \Delta S_{\text{CG}}(R_{IJ}) \rangle$ giving $S_{\text{map}}^{(2)}(R_{IJ}) = \Delta S_{\text{CG}}(R_{IJ})$. This relationship clearly demonstrates a link between the missing entropy (due to coarsening) and the entropic contribution to the CG PMF, i.e., the entropy associated with FG configurations that are degenerate in CG configuration space is encoded in the CG PMF. Here, we note that we expect S_{map} to exactly correspond with the entropic contribution to the CG PMF in the limit of perfectly expressive basis functions and infinite sampling. For practical reasons, we instead represent the CG PMF in a pairwise form, and further decompose it into energy and entropy contributions. We then approximate S_{map} on the basis of the pairwise entropy integrated over all CG configurations using eq (2). Therefore, one goal of the present work is to verify the correspondence shown in eq (2) and further utilize this pairwise entropy function to construct a transferable CG force field.

Even if one can recover an entropy from utilizing eqs (1)–(3), a problem still remains: to accurately relate it to the corresponding entropy at the FG resolution (the “missing” entropy) and to understand its physical meaning. We find that the missing entropy term becomes clearer when applied to single-site CG fluid models. In the single-site model, the CG system will only have translational motion, while at the FG resolution various rotational and vibrational motions exist and are coupled with the translational motions. In other words, the single-site CG model can clearly differentiate the mapped entropy (rotational or vibrational motion) with the naïvely observed entropy (only translational motion). Yet, this argument necessitates a procedure to separate the entropies originating from different motions in the system.

Our strategy is to leverage the two-phase thermodynamic (2PT) method⁴³ and to further decompose the entropy into its modal contributions. In 2PT, the partition function is constructed by partitioning the density of states of the system into translational, rotational, and vibrational contributions. Thermodynamic properties are then obtained by applying quantum statistics to the solid-like component and classical statistics to the gas-like component. Due to its simplicity and efficiency, the 2PT model has been widely applied from simple Lennard-Jones fluid⁴³ to liquid states^{44, 45} with complex phenomena.^{46, 47} Notably, the 2PT method can provide very accurate thermodynamic quantities for liquid states even with the use of short trajectories.^{44, 48} By utilizing the 2PT framework, we can distinguish the translational contribution from the rotational and vibrational contributions and relate them to the configurational entropy at the FG resolution (see Fig. 1). However, any other free energy methodology, such as quasiharmonic analyses, that can extract different modal contributions would work under the suggested framework and would not be limited to only the 2PT method.

Overall, simulations are carried out for both FG and CG systems where the CG models are effectively constructed by the Multiscale Coarse-Graining (MS-CG) formalism,^{19,22,49} although alternative bottom-up CG methods could be used. In practice, the CG parameterization is performed by variational force-matching to minimize the force residuals between the FG and CG ensembles (a detailed procedure is described in Section 5–3 in SI). From the N sets of CG PMFs generated from force-matching, $S_{CG}(R)$ is obtained by a finite difference derivative of the pairwise potential at temperature T_l and T_N given by^{28,50}

$$\Delta S_{CG}(R) \approx - \frac{\Delta W_{CG}(R, T_N) - \Delta W_{CG}(R, T_l)}{T_N - T_l}.$$

Finally, we obtain a complete expression for the mapped entropy as a function of the CG variables. Additional smoothing or filtering can also be performed to the differences in the pairwise forces before the integration step in eq (S12). At the target temperature $T = T^*$, the FG entropy and naïve CG entropy is calculated using the 2PT entropy. The missing CG entropy is subsequently recuperated by utilizing the mapped entropy and the configurations of the CG system at $T = T^*$ as an ensemble average of the pairwise mapping entropy function: $S_{\text{map}} = \langle \sum_{IJ} S_{\text{map}}^{(2)}(R_{IJ}) \rangle = \langle \sum_{IJ} \Delta S_{CG}(R_{IJ}) \rangle$.

The current framework has certain limitations from the approximations that we have introduced. First, we assumed that the CG force fields are adequately pairwise decomposable. However, higher-order correlations may not be captured and thus our present approach may not directly work in systems with non-negligible many-body effects. Further improvements of the present algorithm may be feasible by using stable numerical algorithms to calculate the finite difference, e.g., self-consistent basis sets,⁵⁰ and post-processing the numerical noise from the force-matched profile,⁵¹ e.g., using the Bayesian regularization approach.⁵² These numerical issues are not as evident in this work for liquid mixtures, but certain approaches should be considered for complex systems such as lipids or polypeptides.

Representability: Application to bulk liquids

For single-site CG models where the configurational entropy of the CG system originates from translational motion, we hypothesize that the entropic contributions from rotational and vibrational motions are mapped to the CG PMF. It follows that the value of the mapping entropy is dependent on the vibrational and rotational motions at the FG resolution. For example, the mapping entropy should increase as the vibrational or rotational entropies increase. In order to test our hypothesis, we first designed a proof-of-concept study by investigating the pairwise entropy function in different single-site CG neopentane moieties by gradually tuning the length and stiffness of its C-C bonds. We find that the value of the pairwise entropy functions from different CG neopentane systems gradually increases as the C-C bonds become longer or softer (SI Section S2). With this in mind, we now present results for more complex liquid systems: methanol and chloroform molecules. Using the schemes and details that are described above, the pairwise thermodynamic quantities are obtained and plotted in Fig. 2.

As is immediately evident, the pairwise entropy functions (Fig. 2a) demonstrate approximately positive definite behavior and vanish at large distances. While the energetic profile of methanol has features that are representative of non-spherical interactions, such as the Gay-Berne interaction,⁵³ the profile of chloroform is closer to spherical interactions, which is evident from the small separation between the two potential minima. These features will be further addressed in the next section.

Using the pairwise entropy function $S_{\text{map}}^{(2)}(R)$ from Fig. 2a, the global mapped entropy of the methanol and chloroform systems is readily calculated and shown in Fig. 3. Here, we introduce the radial distribution function $g(R)$ and a change of variables to calculate the radial mapped entropies and integrate them to obtain the overall mapped entropies:

$$S_{\text{map}} = \rho \int dR \cdot 4\pi R^2 g(R) \cdot S_{\text{map}}^{(2)}(R) \quad (4)$$

This approach gives consistent values to other numerical techniques, such as averaging the quantity $-\langle W_{\text{CG}}(R_{II}) \rangle / T$ over the ensemble (SI Section 3). In Fig. 3a, the major peaks with non-negligible intensities occur in regions with large $g(R)$ values, but soon the radial

entropy $S_{\text{rad}}(R) = 4\pi R^2 \rho g(R) \cdot S_{\text{map}}^{(2)}(R)$ decays to zero because $\lim_{R \rightarrow R_{\text{cut}}} S_{\text{map}}^{(2)}(R) = 0$. To check

convergence of the mapped entropy, we integrate $S_{\text{rad}}(R)$ from $R = 0$ to $R = R_{\text{cut}}$ by varying R_{cut} from 3 to 20 Å. The changes to the overall mapped entropy are depicted in Fig. 3b, demonstrating that the value quickly converges beyond the first coordination shell. It is worth noting that the current approach is mainly suitable for systems without strong long-range correlations due to the contribution from $R^2 g(R) \cdot S_{\text{map}}^{(2)}(R)$.

To establish a connection between the CG entropy and its correspondence at the FG resolution, we first computed the all-atom configurational entropies of methanol and chloroform at $T = 300$ K by utilizing the 2PT method.⁴³ Here, liquid systems were modeled using the Optimized Potential for Liquid Simulations (OPLS) force field as listed in Table 1.⁵⁴ The all-atom entropies are first validated by comparison to previously reported entropy values using same method⁴⁴ and the OPLS force field. Slight differences observed for the methanol system likely arise due to differences in simulated temperatures, resulting in consistently larger entropy values than the reported values (Ref 44 used $T = 298$ K). For chloroform, we used a modified version of the OPLS force field from Ref 55, but the general trends remain consistent.

For the CG systems, the naïve CG entropy similarly calculated from the 2PT method is shown in Table 1, which corresponds to the translational entropy given the absence of intramolecular vibrations and rotations due to our CG mapping. Likewise, we approximate the entropy that is folded into the CG PMF from Fig. 3b, which corresponds to the rotational and vibrational entropies that are lost due to the CG mapping. In comparing the all-atom and CG translational entropies, we note larger values in the latter case, which likely arise due to the following factor. First, we restrict ourselves to a pairwise basis set²² to represent the effective CG potential, which may not capture the complete many-body PMF; for instance, single-site methanol is inherently non-spherical at the atomistic level, which can only be approximated to a certain extent with simple pairwise interactions (see Fig. 2). More importantly, we also find that our approximations of the mapped (i.e., missing) CG entropies for both systems recapitulates reasonably well the rotational and vibrational contributions at the FG resolution, even though the CG model for methanol somewhat overestimates the mapping entropy contribution. This analysis therefore suggests that our approach can faithfully represent the entropy of CG systems in a pairwise decomposable manner by explicitly including the missing entropy, in which the latter quantity corresponds to the FG entropy that is lost from CG mapping.

Transferability: Temperature transferability of bulk liquids

As mentioned previously, the transferability problem arises due to difficulty in capturing the many-body PMF, especially since the PMF will change in unknown ways for new systems given its state-point dependency.^{14, 15, 24} In particular, one must consider the non-trivial contribution of the CG entropy to the PMF, which has been challenging to unambiguously quantify. An end goal of this work is to enable a general CG potential that retains high-fidelity in different molecular environments and thermodynamic state points.¹⁴ We will

focus on two different cases: (i) different temperatures and (ii) different systems or chemical environments, i.e., bulk liquid and liquid mixtures. As a first step in constructing transferable CG interactions for a given system at different state points, the most direct approach is to interpolate the PMFs at different temperatures by leveraging the energy-entropy decomposition expressed in eq (3) within valid temperature ranges, i.e., within a given phase. In practice, we estimate the pairwise terms in the CG PMF at different temperatures, where the pairwise thermodynamic functions $U_{CG}(R)$ and $S_{CG}(R)$ are from Fig. 2 and assumed here to be temperature independent (i.e., a simple linear dependence on temperature of $W_{CG}(R; T)$).

We find that the analytically-fitted effective CG interactions (using eq (3)) for methanol and chloroform are in excellent agreement with respect to the interactions derived from MS-CG at each state-point, as shown in Fig. 4. A nearly linear dependence of the CG PMF on temperature is consistent with previous reports.^{28, 56, 57} It is worthwhile to note that we used the following ranges to obtain the pairwise entropy functions of each system: 250–350 K for methanol and 250–325 K for chloroform. However, in Fig. 4, we not only interpolated the PMFs at temperatures within the training range but also extrapolated them to even higher temperatures outside of our training set. For methanol, we find that the range of extrapolation can be much larger than those reported in previous works (330 K for Ref 28 and 300 K for Ref 56). Remarkably, the extrapolated CG interactions can also reproduce the effective PMF correctly, even in the long-range regimes (Fig. 4b and 4d), while assuming only linear temperature dependence: $W_{CG}(R; T) = U_{CG}(R) - T S_{CG}(R)$. This success in extrapolating the CG PMF suggests that we can extend this approach and construct “phase-transferable” CG models by combining pairwise energy and entropy functions from different phases. We are, for example, currently pursuing the development of CG force fields that are transferable between liquid and vapor phases for a future publication.

Transferability to different chemical environments: “Liquid mixtures”

In order to transfer the effective CG interactions to different chemical environments, we construct liquid mixtures composed of methanol and chloroform. Constructing bottom-up CG models that are transferable from bulk to mixed-liquid states has been recognized as a difficult problem, and has therefore been the focus of recent research efforts.^{30, 32} However, recently suggested methods remain system-dependent, thereby limiting their applicability. However, as is common practice in all-atom and top-down models, a “combining rule” for arbitrary CG models is one potential solution. Even though combining (or mixing) different PMFs may seem nearly impossible due to their complex nature, we seek to predict effective CG interactions between two different molecules by solely utilizing information derived from bulk states.

To address the aforementioned challenges, we show that the energy-entropy decomposition of bulk state PMFs can be utilized to predict the PMFs associated with single-site CG liquid mixtures, especially cross-interactions, in a systematic way. Rather than combining the PMF naively, our framework discriminates and combines the energetic and entropic terms separately. Finally, the interpolated (or mixed) PMF of the system is readily obtained by adding these two functions. To more effectively predict the interaction between methanol

and chloroform CG beads, we approximate the methanol CG bead as a spherocylindrical particle and the chloroform CG bead as a spherical particle. These choices were motivated by the energetic interaction profiles calculated in Fig. 2b, which suggest that methanol has a rod-like interaction⁵⁸ while chloroform retains spherical symmetry with a small degree of anisotropy.³⁰ By adopting this assumption, the methanol-chloroform cross-interaction was simplified to a rod-sphere interaction. Based on previous statistical mechanical developments of rod-sphere interactions using Gay-Berne and Lennard-Jones functional forms^{59, 60}, we first extracted effective Gay-Berne^{53, 58} and Lennard-Jones (LJ 6-4)⁶¹ parameters from the self-interaction PMFs and projected the cross-interaction onto an approximated pairwise form (please refer to the SI Section 6). As a result, we then obtained a simplified expression of the rod-sphere energetic interaction term as:

$$\Delta U_{\text{MeOH-CCl}_3\text{H}}(R) \approx \frac{27}{4} \sqrt{\epsilon_{\text{MeOH}} \cdot \epsilon_{\text{CCl}_3\text{H}}} \left[\left(\frac{\sigma_{\text{CCl}_3\text{H}}}{R} \right)^6 - \left(\frac{\sigma_{\text{CCl}_3\text{H}}}{R} \right)^4 \right] \quad (5)$$

with $\sigma_{\text{CCl}_3\text{H}} = \sigma_{r-s} = 3.1893 \text{ \AA}$ and $\epsilon_{r-s} = \sqrt{\epsilon_{\text{MeOH}} \cdot \epsilon_{\text{CCl}_3\text{H}}} = 0.9503 \text{ kcal/mol}$. The pairwise self-interactions (methanol-methanol, chloroform-chloroform) were modeled using the procedure detailed in the previous section, with additional details in the SI.

While mixing the energetic interactions is similar to conventional combining rules for interaction parameters, mixing the entropic functions is relatively different. Specifically, we point out that the entropic mixing discussed in this work is a pairwise quantity that is a function of pair distance (i.e., not the conventional entropy of mixing), which, to our knowledge, is seldomly explored. Here, we suggest a combining rule for pairwise entropy functions. Due to the near-positive definite nature of the obtained entropy function from the single-site CG mapping, we take a geometric average of the entropy functions at a given distance R . This expression is given by $\Delta S_{\text{MeOH-CCl}_3\text{H}}^{\text{mix}}(R) = \sqrt{(\Delta S_{\text{MeOH}}^{\text{bulk}}(R)) \cdot (\Delta S_{\text{CCl}_3\text{H}}^{\text{bulk}}(R))}$ (Note that $S_{\text{CG}}(R) > 0$). In turn, we note that such an entropic combining rule is consistent with the conventional Lorentz-Berthelot mixing rules^{62, 63} (see the SI). From the proposed mixing protocol, we can compute the pairwise energy and entropy functions for the methanol-chloroform interaction and compare them to the actual MS-CG interactions from force-matching the mixture system (Fig. 5a). Notably, excellent agreement is observed for both the energetic and entropic contributions in Fig. 5a, which suggests that the interpolated model can be transferred to other temperatures as well. We next perform CG simulations using the fitted CG interactions to evaluate the fidelity of the mixed CG model.

To evaluate the structural correlations of methanol-chloroform pairs, we calculated the radial distribution function (RDF), or $g(r)$, of methanol-chloroform and compared it to the exact CG mapped all-atom reference as shown in Fig. 5b. The RDF of the interpolated CG model is smoother than the mapped all-atom reference due to the use of simple CG energetic functions for combining. In particular, it deviates in the first peak width that can be interpreted as smoothing over underlying structure that appears in the all-atom structure. Nonetheless, the structural agreement is quite reasonable (the peak value near 4.5 Å is

almost identical), given the fact that the interpolated CG PMF was constructed without any prior information of the exact CG cross-interaction.

In order to investigate higher-order correlations, especially given the inhomogeneous nature of this system,⁶⁴ we perform cluster analysis of methanol and chloroform clusters based on a previously reported protocol that utilizes connected graphs within the first coordination shell to identify clusters.²⁹ Surprisingly, we find that the cluster distribution of the interpolated CG models and the CG mapped atomistic system matches almost perfectly in both methanol and chloroform regardless of the observation scheme (averaged observation in Fig. 5c and logarithm of overall clusters in Fig. 5d). Altogether, both thermodynamic and structural properties derived from our proposed strategy to construct transferable CG interactions appear to faithfully reproduce the actual CG properties of the mixture as directly obtained by force-matching to the mixture.

To summarize, our work elucidates the nature of CG entropy as the mapped (missing) entropy from the many-body CG PMF determined by the CG mapping. Using single-site CG models, we clarify the implicit embedding of vibrational and rotational entropies within the CG PMF, while only the translational entropy is captured by naïve arguments. Furthermore, we demonstrate the use of a free energy decomposition framework to impart transferable bottom-up CG force fields for non-parameterized state points. Our findings in the methanol/chloroform mixture system suggest a high-fidelity combining rule to approximate CG PMFs for different chemical environments by separately mixing the energetic and entropic components of the bulk-phase CG PMFs.

In order to combine CG PMFs, an entropic combining rule is suggested based on mixing rules that are introduced in this work. Nevertheless, it is evident that much of the physical nature of the pairwise entropy function and its combining rules remains unexplored, suggesting that the proposed framework will have various potential applications and future directions. A natural extension of the proposed combining rule for simple liquids would be to model complex interacting systems that necessitate CG modeling: protein-protein interactions or proteins in water. For example, we plan to integrate the current approach with a recently proposed bottom-up CG modeling framework for complex biomolecules in water (e.g., lipids), which represent solvent effects through the use of virtual sites.⁶⁵ After successfully applying the approach to more complex biomolecules, the next step will be to investigate the correct combining rules for pairwise thermodynamic quantities in such systems. We also note that the pairwise entropy function obtained from different temperatures might be useful to compare and validate different CG models. In contrast to more *ad hoc* top-down CG models, the present framework is expected to facilitate bottom-up CG modeling by providing a systematic procedure to transfer CG force fields to different thermodynamic state points and compositions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

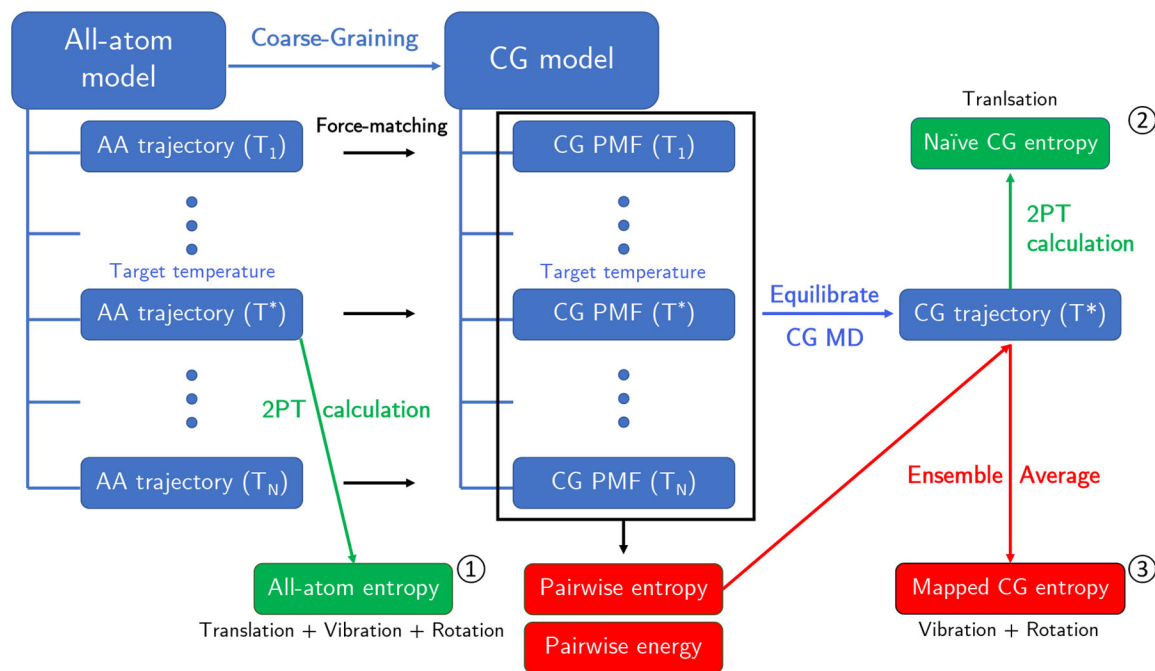
This material is based upon work supported by the National Science Foundation (NSF Grant CHE-1465248). Simulations were performed using computing resources provided by the University of Chicago Research Computing Center (RCC). J.J. acknowledges a graduate fellowship from the Kwanjeong Educational Foundation. A.J.P. acknowledges support from the Ruth L. Kirschstein National Research Service Award Postdoctoral Fellowship from the National Institutes of Health (F32-AI150477). We gratefully acknowledge the Samsung HumanTech Award and Yining Han for stimulating conversations on this project.

REFERENCES

1. MacKerell AD Jr; Bashford D; Bellott M; Dunbrack RL Jr; Evanseck JD; Field MJ; Fischer S; Gao J; Guo H; Ha S All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 1998, 102, 3586–3616. [PubMed: 24889800]
2. Karplus M; McCammon JA Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol* 2002, 9, 646. [PubMed: 12198485]
3. Adcock SA; McCammon JA Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev* 2006, 106, 1589–1615. [PubMed: 16683746]
4. Shaw DE; Maragakis P; Lindorff-Larsen K; Piana S; Dror RO; Eastwood MP; Bank JA; Jumper JM; Salmon JK; Shan Y Atomic-level characterization of the structural dynamics of proteins. *Science* 2010, 330, 341–346. [PubMed: 20947758]
5. Schlick T; Collepardo-Guevara R; Halvorsen LA; Jung S; Xiao X Biomolecular modeling and simulation: a field coming of age. *Q. Rev. Biophys* 2011, 44, 191–228. [PubMed: 21226976]
6. Frenkel D; Smit B Understanding molecular simulation: from algorithms to applications. Elsevier: 2001; Vol. 1.
7. Allen MP; Tildesley DJ Computer simulation of liquids. Oxford university press: 2017.
8. Müller-Plathe F Coarse-graining in polymer simulation: from the atomistic to the mesoscopic scale and back. *ChemPhysChem* 2002, 3, 754–769.
9. Scheraga HA; Khalili M; Liwo A Protein-folding dynamics: overview of molecular simulation techniques. *Annu. Rev. Phys. Chem* 2007, 58, 57–83. [PubMed: 17034338]
10. Voth GA Coarse-graining of condensed phase and biomolecular systems. CRC press: 2008.
11. Peter C; Kremer K Multiscale simulation of soft matter systems-from the atomistic to the coarse-grained level and back. *Soft Matter* 2009, 5, 4357–4366.
12. Murtola T; Bunker A; Vattulainen I; Deserno M; Karttunen M Multiscale modeling of emergent materials: biological and soft matter. *Phys. Chem. Chem. Phys* 2009, 11, 18691892.
13. Riniker S; Allison JR; van Gunsteren WF On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys* 2012, 14, 12423–12430. [PubMed: 22678152]
14. Noid W Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys* 2013, 139, 090901. [PubMed: 24028092]
15. Saunders MG; Voth GA Coarse-graining methods for computational biology. *Annu. Rev. Biophys* 2013, 42, 73–93. [PubMed: 23451897]
16. Henderson R A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A* 1974, 49, 197–198.
17. Noid W Systematic methods for structurally consistent coarse-grained models. In *Biomolecular Simulations*, Springer: 2013; pp 487–531.
18. Wagner JW; Dama JF; Durumeric AE; Voth GA On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys* 2016, 145, 044108. [PubMed: 27475349]
19. Izvekov S; Voth GA A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* 2005, 109, 2469–2473. [PubMed: 16851243]
20. Izvekov S; Voth GA Multiscale coarse graining of liquid-state systems. *J. Chem. Phys* 2005, 123, 134105. [PubMed: 16223273]
21. Noid W; Chu J-W; Ayton GS; Krishna V; Izvekov S; Voth GA; Das A; Andersen HC The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys* 2008, 128, 244114. [PubMed: 18601324]

22. Noid W; Liu P; Wang Y; Chu J-W; Ayton GS; Izvekov S; Andersen HC; Voth GA The multiscale coarse-graining method. II. Numerical implementation for coarsegrained molecular models. *J. Chem. Phys* 2008, 128, 244115. [PubMed: 18601325]
23. Allen EC; Rutledge GC Evaluating the transferability of coarse-grained, density-dependent implicit solvent models to mixtures and chains. *J. Chem. Phys* 2009, 130, 034904. [PubMed: 19173540]
24. Dunn NJ; Foley TT; Noid WG Van der Waals perspective on coarse-graining: Progress toward solving representability and transferability problems. *Acc. Chem. Res* 2016, 49, 2832–2840. [PubMed: 27993007]
25. Carbone P; Varzaneh HAK; Chen X; Müller-Plathe F Transferability of coarsegrained force fields: The polymer case. *J. Chem. Phys* 2008, 128, 064904. [PubMed: 18282071]
26. Mullinax J; Noid W Extended ensemble approach for deriving transferable coarsegrained potentials. *J. Chem. Phys* 2009, 131, 104110.
27. Krishna V; Noid WG; Voth GA The multiscale coarse-graining method. IV. Transferring coarse-grained potentials between temperatures. *J. Chem. Phys* 2009, 131, 024103. [PubMed: 19603966]
28. Lu L; Voth GA The multiscale coarse-graining method. VII. Free energy decomposition of coarse-grained effective potentials. *J. Chem. Phys* 2011, 134, 06B607.
29. Dama JF; Jin J; Voth GA The theory of Ultra-Coarse-Graining. 3. Coarse-grained sites with rapid local equilibrium of internal states. *J. Chem. Theory Comput* 2017, 13, 10101022.
30. Jin J; Voth GA Ultra-Coarse-Grained Models Allow for an Accurate and Transferable Treatment of Interfacial Systems. *J. Chem. Theory Comput* 2018, 14, 2180–2197. [PubMed: 29481754]
31. Rosenberger D; van der Vegt NF Addressing the temperature transferability of structure based coarse graining models. *Phys. Chem. Chem. Phys* 2018, 20, 6617–6628. [PubMed: 29451286]
32. Sanyal T; Shell MS 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. [PubMed: 29466859]
33. Louis A Beware of density dependent pair potentials. *J. Phys.: Condens. Matter* 2002, 14, 9187.
34. Johnson ME; Head-Gordon T; Louis AA Representability problems for coarsegrained water potentials. *J. Chem. Phys* 2007, 126, 144509. [PubMed: 17444725]
35. 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: Soft Matter Biol. Phys* 2009, 28, 221–229.
36. 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. [PubMed: 20158022]
37. Guenza M Thermodynamic consistency and other challenges in coarse-graining models. *Eur. Phys. J.: Spec. Top* 2015, 224, 2177–2191.
38. Foley TT; Shell MS; Noid W The impact of resolution upon entropy and information in coarse-grained models. *J. Chem. Phys* 2015, 143, 12B601_1.
39. Noid W; Chu J-W; Ayton GS; Voth GA Multiscale coarse-graining and structural correlations: Connections to liquid-state theory. *J. Phys. Chem. B* 2007, 111, 4116–4127. [PubMed: 17394308]
40. Izvekov S Towards an understanding of many-particle effects in hydrophobic association in methane solutions. *J. Chem. Phys* 2011, 134, 034104. [PubMed: 21261327]
41. Cao Z; Dama JF; Lu L; Voth GA Solvent free ionic solution models from multiscale coarse-graining. *J. Chem. Theory Comput* 2012, 9, 172–178. [PubMed: 26589020]
42. 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. [PubMed: 26574265]
43. Lin S-T; Blanco M; Goddard III WA The two-phase model for calculating thermodynamic properties of liquids from molecular dynamics: Validation for the phase diagram of Lennard-Jones fluids. *J. Chem. Phys* 2003, 119, 11792–11805.
44. Pascal TA; Lin S-T; Goddard III WA Thermodynamics of liquids: standard molar entropies and heat capacities of common solvents from 2PT molecular dynamics. *Phys. Chem. Chem. Phys* 2011, 13, 169–181. [PubMed: 21103600]
45. Pascal TA; Goddard III WA Interfacial thermodynamics of water and six other liquid solvents. *J. Phys. Chem. B* 2014, 118, 5943–5956. [PubMed: 24820859]

46. Pascal TA; Goddard WA; Jung Y Entropy and the driving force for the filling of carbon nanotubes with water. *Proc. Natl. Acad. Sci. U. S. A* 2011, 108, 11794–11798. [PubMed: 21709268]
47. Jin J; Goddard III WA Mechanisms underlying the Mpemba effect in water from molecular dynamics simulations. *J. Phys. Chem. C* 2015, 119, 2622–2629.
48. Lin S-T; Maiti PK; Goddard III WA Two-phase thermodynamic model for efficient and accurate absolute entropy of water from molecular dynamics simulations. *J. Phys. Chem. B* 2010, 114, 8191–8198. [PubMed: 20504009]
49. Lu L; Izvekov S; Das A; Andersen HC; Voth GA Efficient, regularized, and scalable algorithms for multiscale coarse-graining. *J. Chem. Theory Comput.* 2010, 6, 954–965. [PubMed: 26613319]
50. Wagner JW; Dama JF; Voth GA Predicting the sensitivity of multiscale coarsegrained models to their underlying fine-grained model parameters. *J. Chem. Theory Comput* 2015, 11, 3547–3560. [PubMed: 26574440]
51. Das A; Lu L; Andersen HC; Voth GA The multiscale coarse-graining method. X. Improved algorithms for constructing coarse-grained potentials for molecular systems. *J. Chem. Phys* 2012, 136, 194115. [PubMed: 22612088]
52. Liu P; Shi Q; Daumé III H; Voth GA A Bayesian statistics approach to multiscale coarse graining. *J. Chem. Phys* 2008, 129, 12B605.
53. Gay J; Berne B Modification of the overlap potential to mimic a linear site-site potential. *J. Chem. Phys* 1981, 74, 3316–3319.
54. Jorgensen WL; Maxwell DS; Tirado-Rives J Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc* 1996, 118, 11225–11236.
55. Dodda LS; Vilseck JZ; Tirado-Rives J; Jorgensen WL 1.14* CM1A-LBCC: localized bond-charge corrected CM1A charges for condensed-phase simulations. *J. Phys. Chem. B* 2017, 121, 3864–3870. [PubMed: 28224794]
56. Lebold KM; Noid WG Systematic study of temperature and density variations in effective potentials for coarse-grained models of molecular liquids. *J. Chem. Phys* 2019, 150, 014104. [PubMed: 30621403]
57. Lebold KM; Noid W Dual approach for effective potentials that accurately model structure and energetics. *J. Chem. Phys* 2019, 150, 234107. [PubMed: 31228924]
58. Golubkov PA; Ren P Generalized coarse-grained model based on point multipole and Gay-Berne potentials. *J. Chem. Phys* 2006, 125, 064103.
59. Antypov D; Cleaver DJ The role of attractive interactions in rod–sphere mixtures. *J. Chem. Phys* 2004, 120, 10307–10316. [PubMed: 15268056]
60. Cleaver DJ; Care CM; Allen MP; Neal MP Extension and generalization of the Gay-Berne potential. *Phys. Rev. E* 1996, 54, 559.
61. Fabiola F; Bertram R; Korostelev A; Chapman MS An improved hydrogen bond potential: impact on medium resolution protein structures. *Protein Science* 2002, 11, 1415–1423. [PubMed: 12021440]
62. Lorentz H Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys* 1881, 248, 127–136.
63. Berthelot D Sur le mélange des gaz. *C. R. Hebd. Séances Acad. Sci* 1898, 126, 1703.
64. Tribello GA; Giberti F; Sosso GC; Salvalaglio M; Parrinello M Analyzing and driving cluster formation in atomistic simulations. *J. Chem. Theory Comput* 2017, 13, 13171327.
65. Pak AJ; Dannenhoffer-Lafage T; Madsen JJ; Voth GA Systematic CoarseGrained Lipid Force-Fields with Semi-Explicit Solvation via Virtual Sites. *J. Chem. Theory Comput* 2019.



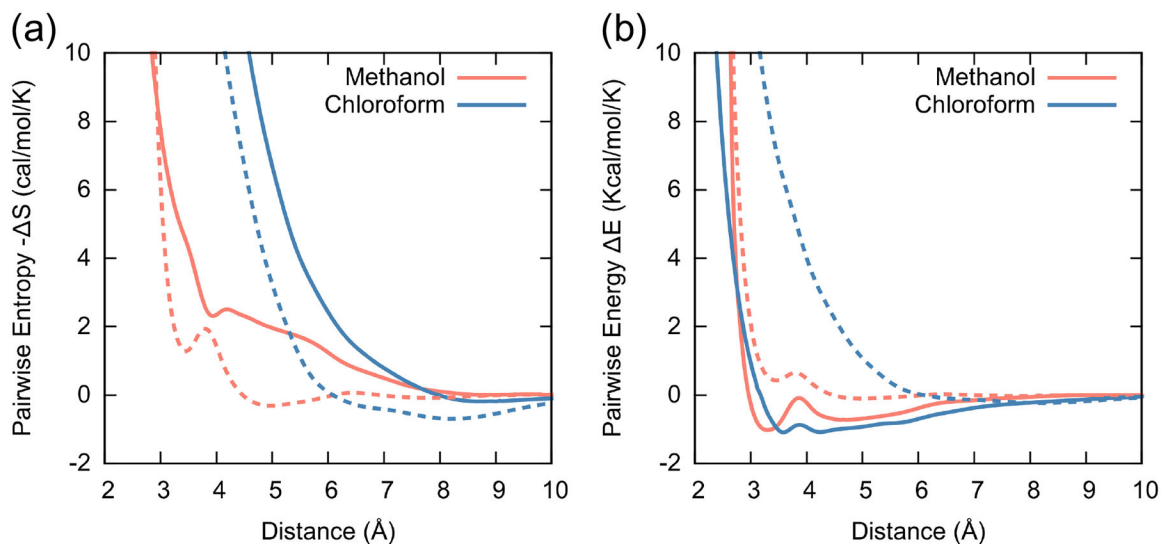
Green box denotes the Gibbs entropy

$$\text{CG PMF}(T^*) = \text{Pairwise energy} - T^* \times \text{Pairwise entropy}$$

$$\text{All-atom entropy}^{(1)} = \text{Naïve CG entropy}^{(2)} + \text{Mapped CG entropy}^{(3)}$$

Figure 1:

Schematic diagram describing the procedure for calculating different entropy quantities in all-atom and CG systems. The entropy of a single-site CG liquid system consists of two components: the naïve (translation) and mapped CG entropies (vibration and rotation). Both all atom entropies and naïve CG entropies (based on the Gibbs definition) are directly obtained by 2PT calculations using MD trajectories. The mapped CG entropy is calculated separately using the pairwise entropy function from the CG PMF using eq (4). The representability relationship for entropy (bottom) asserts that the all-atom entropy should be identical to the sum of naïve and mapped CG entropies.

**Figure 2:**

Pairwise thermodynamic quantities for liquids systems: methanol (red) and chloroform (blue). (a) The pairwise mapping entropy function $S_{\text{map}}^{(2)}(R)$ (solid lines) calculated from the scheme depicted in Fig. 1. (b) The pairwise energy function (solid lines) calculated from differences between the PMF and the pairwise entropy function. For comparison, the pairwise PMFs of both liquids are shown as dashed lines in each figure. For (a), the PMF values are additionally scaled by a factor of $1/T$ with $T = 300$ K.

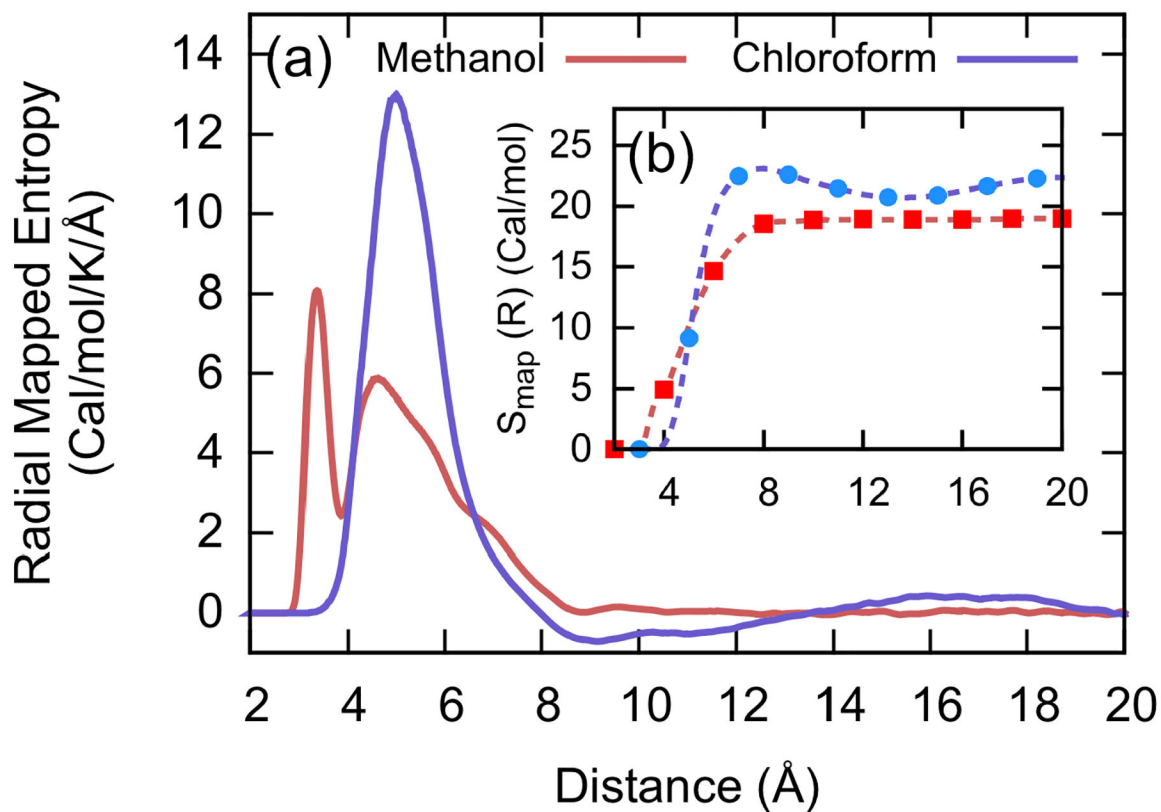


Figure 3:

Mapped entropy for methanol (red) and chloroform (purple) system using the pairwise entropy function. (a) The radial mapped entropy component $S_{rad}(R) = 4\pi R^2 \rho g(R) \cdot S_{map}^{(2)}(R)$.

(b) The global mapped CG entropy obtained by integrating the radial mapped entropy with different cutoff radii for methanol (red, square) and chloroform (blue, circle).

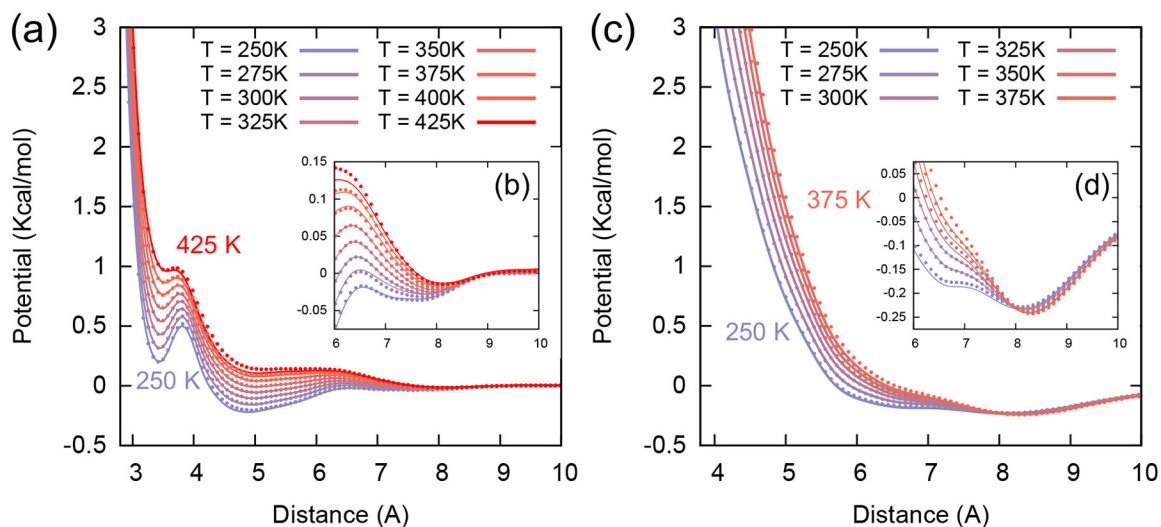
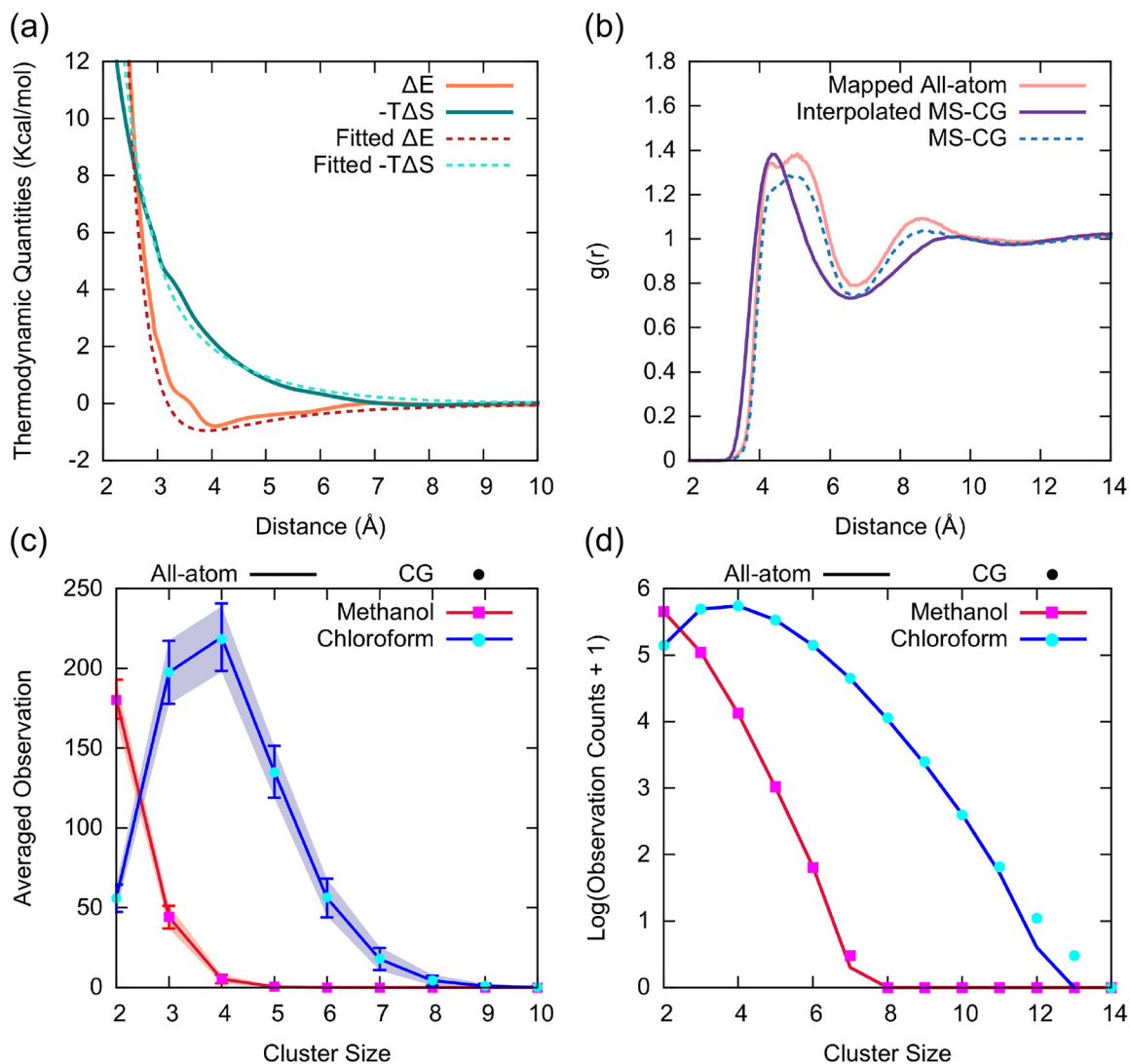


Figure 4: Effective CG interactions of the methanol and chloroform systems for different temperatures by comparing the fitted PMFs (dots) with the conventional MS-CG PMFs (lines). (a) Methanol-Methanol CG PMFs at selected temperatures from 250 K to 425 K. (b) Magnified CG PMFs of methanol at the long-range region between 6 and 10 Å. (c) Chloroform-Chloroform CG PMFs at selected temperatures from 250 K to 375 K. (d) Magnified CG PMFs of chloroform at the long-range region between 6 and 10 Å.

**Figure 5:**

Assessment of the transferability of CG models in the liquid mixture utilizing a proposed combining rule based on energy-entropy decomposition. Here, each liquid molecule was mapped to single CG bead, (a) Thermodynamic quantities of the methanol-chloroform CG cross interaction: energetic (orange) and entropic (green) interactions from the proposed combining rule (dotted lines) comparing to the actual MS-CG (solid lines), (b) Methanol-chloroform $g(r)$ for CG mapped atomistic and interpolated MS-CG simulations. The $g(r)$ function obtained from the actual MS-CG model for the CG mixture is also depicted (dotted lines) to provide the upper limit of possible CG models. Histograms of observed methanol (red, magenta) and chloroform (blue, cyan) cluster sizes in mapped all-atom and interpolated CG models in (c) natural (averaged) scales and (d) log scales.

Table 1:

Comparison of average molar entropy for methanol and chloroform with different contributions: Molar entropies calculated at all-atom resolution, naïve CG entropies (translation only) from the 2PT calculation, and mapped (or missing) CG entropy from Fig. 3b (due to rotation and vibration).

Entropy (Cal mol ⁻¹ HK ⁻¹)	All-atom	MS-CG		Reference 2PT ⁴⁴	
		<i>Naïve</i>	<i>Missing</i>		
Methanol					
Translation	16.12		21.11		16.03
Rotation	11.53		0.00		11.38
Vibration	1.73	13.26	0.00	18.95	13.09
Chloroform					
Translation	20.32		24.96		21.12
Rotation	18.95		0.00		19.20
Vibration	5.67	24.62	0.00	22.33	24.85