Supporting Information for

"Understanding Missing Entropy in Coarse-Grained Systems: 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

*Corresponding author: gavoth@uchicago.edu

S1. Derivations of the CG entropy expression

This section will provide a brief proof of eq 1 in the main text based on ref 1 and 2. The first derivation of the CG entropy expression is shown in ref 2, but in this section we will follow the proof given in ref 1. As defined in the manuscript, the CG configurational entropy is denoted as $S_{\mathrm{CG}}(\omega_{\mathrm{CG}})$ which can be expressed as the following:

$$S_{\mathrm{CG}}(\omega_{\mathrm{CG}}) = -k_B \int_{\omega_{\mathrm{FG}}} d\mathbf{r}^{\mathrm{n}} p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}}) \ln p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}}). \tag{S1}$$

In the above equation, we ensure that the integral on the right-hand side is expressed over $\mathbf{r}^{\mathrm{n}} \in \omega_{\mathrm{FG}}$. This integral is therefore separated by the FG and CG parts, respectively, using $\int_{\omega_{\mathrm{CG}}} d\mathbf{R}^{\mathrm{N}} \int d\mathbf{r}^{\mathrm{n}} \delta\left(\mathbf{R}^{\mathrm{N}} - M(\mathbf{r}^{\mathrm{n}})\right)$,

$$S_{\mathrm{CG}}(\omega_{\mathrm{CG}}) = -k_B \int_{\omega_{\mathrm{CG}}} d\mathbf{R}^{\mathrm{N}} p_{\omega_{\mathrm{CG}}}(\mathbf{R}^{\mathrm{N}}) \left(\ln p_{\omega_{\mathrm{CG}}}(\mathbf{R}^{\mathrm{N}}) + \int d\mathbf{r}^{\mathrm{n}} \delta\left(\mathbf{R}^{\mathrm{N}} - M(\mathbf{r}^{\mathrm{n}})\right) \frac{p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}})}{p_{\omega_{\mathrm{CG}}}(\mathbf{R}^{\mathrm{N}})} \ln \frac{p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}})}{p_{\omega_{\mathrm{CG}}}(\mathbf{R}^{\mathrm{N}})} \right). \tag{S2}$$

The first term is the naïve CG entropy $S_{\mathrm{CG}}^{\mathrm{naïve}}(\omega_{\mathrm{CG}}) = -k_B \int_{\omega_{\mathrm{CG}}} d\mathbf{R}^{\mathrm{N}} p_{\omega_{\mathrm{CG}}}(\mathbf{R}^{\mathrm{N}}) \ln p_{\omega_{\mathrm{CG}}}(\mathbf{R}^{\mathrm{N}})$ while the other term is excess (or missing) entropy $S_{\mathrm{CG}}^{\mathrm{missing}}(\omega_{\mathrm{CG}})$, which can be formulated as the conditional probability $\frac{p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}})}{p_{\omega_{\mathrm{CG}}}(\mathbf{R}^{\mathrm{N}})} = p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}}|\mathbf{R}^{\mathrm{N}})$. This gives

$$\begin{aligned} S_{\mathrm{CG}}^{\mathrm{missing}}(\omega_{\mathrm{CG}}) &= -k_B \int_{\omega_{\mathrm{CG}}} d\mathbf{r}^{\mathrm{n}} \delta\left(\mathbf{R}^{\mathrm{N}} - M(\mathbf{r}^{\mathrm{n}})\right) p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}}|\mathbf{R}^{\mathrm{N}}) \ln p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}}|\mathbf{R}^{\mathrm{N}}) \\ &= \langle -k_B \ln p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}}|\mathbf{R}^{\mathrm{N}}) \rangle_{\omega_{\mathrm{CG}}}. \end{aligned} \tag{S3}$$

Finally, from thermodynamic relationships, we recover eq 1 as seen below:

$$\langle -k_B \ln p_{\omega_{\mathrm{FG}}}(\mathbf{r}^{\mathrm{n}}|\mathbf{R}^{\mathrm{N}}) \rangle_{\omega_{\mathrm{CG}}} = -\frac{1}{T}\left(W_{\mathrm{CG}}(\mathbf{R}^{\mathrm{N}}, T) - \frac{d\beta W_{\mathrm{CG}}(\mathbf{R}^{\mathrm{N}}, T)}{d\beta} \right) = -\frac{dW_{\mathrm{CG}}(\mathbf{R}^{\mathrm{N}})}{dT}. \tag{S4}$$

Here, we denote the $W_{\mathrm{CG}}(\mathbf{R}^{\mathrm{N}})$ as the many-body potential of mean force in the canonical ensemble (constant NVT), as described the main text.

S2. Proof-Of-Concept Study

S2-1. Details of All-atom Simulation

The template neopentane molecule is generated from an automatic topology builder.[3, 4] The overall system is composed of 1,000 neopentane molecules in a cubic box using periodic boundary conditions. The detailed all-atom simulation protocol is described in section S5-1. In the case of neopentane, we used T = 275 K to maintain the system in the liquid phase. From the all-atom trajectories, we coarse-grained the atomistic configurations by mapping each methyl group to one CG bead. Namely, the neopentane molecule was mapped to a 5-site CG model.

S2-2. System Setting

In order to examine the contribution of rotational and vibrational modes on the pairwise entropy, we selectively tuned the C-C bond parameters for neopentane. From the OPLS-AA force field,[5] the interaction coefficients for the harmonic C-C bond is $R = 1.529$ Å and $k = 268$ kcal/mol. We both lengthened and weakened the C-C bond by modifying $R' = 2.0$ Å and $k' = 5$ kcal/mol. We envisage that the longer and weaker C-C bond results in more rotational and vibrational motions of the system. We constructed two different models with tuned bonds: neo-*1* (with one tuned bond) and neo-*4* (all C-C bonds are tuned). In order to obtain the pairwise entropy profile, we performed other atomistic simulations at 245, 260, 290, 305 K with ΔT = 15 K for each system. Finally, the pairwise entropy function $\Delta S_{\mathrm{CG}}(R)$ was obtained by eq 2 and S12 from the main text. Calculated $\Delta S_{\mathrm{CG}}(R)$ functions from neo (default neopentane), neo-*1*, and neo-*4* are shown in Fig. S1 below.

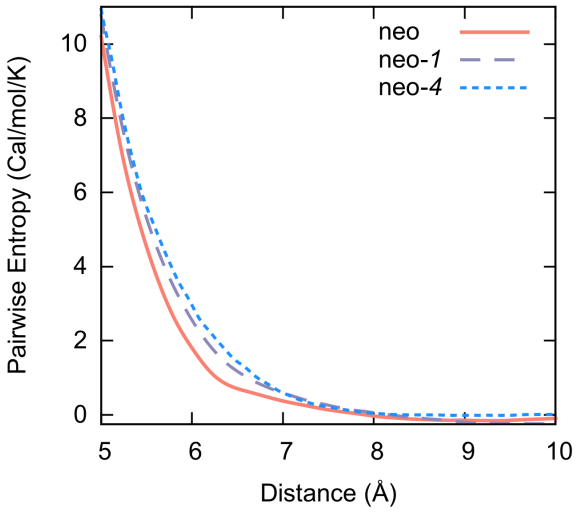

Figure S1: Schematic diagram and changes to pairwise entropy $S^{(2)}_{\mathrm{map}}(R) = \Delta S_{\mathrm{CG}}(R)$ values of the neopentane moieties: default neopentane (neo with red lines), neopentane with one tuned C-C bond (neo-*1* with purple dashes), and neopentane with all tuned C-C bonds (neo-*4* with blue dots).

S2-3. Discussion

From Fig. S1, it is immediately apparent that tuning the C-C bond affects the pairwise entropy function. The regular CG neopentane molecule shows a profile with non-negative values, which further increases after tuning the C-C bonds to be longer and weaker. More importantly, the trend of increasing entropy is consistent from default to neo-*1* to neo-*4*, as expected. While the increment in value may not seem large, the integrated entropy S_{map}, which is proportional to $\int 4\pi R^2 g(R)$, is a non-negligible value.

S3. Calculating the mapping entropy: Possible alternative methods and convergence

S3-1. Other possible schemes to calculate the mapping entropy

We know that the mapping entropy at the CG configurations $\mathbf{R}^{\mathrm{N}}$ is approximately expressed as

$$S_{\mathrm{map}} \approx \langle -\sum_{IJ} \frac{d\Delta W_{\mathrm{CG}}(R_{IJ})}{dT} \rangle = \langle \sum_{IJ} S_{\mathrm{map}}^{(2)}(R_{IJ}) \rangle = \langle \sum_{IJ} \Delta S_{\mathrm{CG}}(R_{IJ}) \rangle . \quad \text{(S5)}$$

In the main text, we calculated this quantity using the pair correlation function of the liquid system: $S_{\mathrm{map}} = \rho \int dR \cdot 4\pi R^2 g(R) \cdot S_{\mathrm{map}}^{(2)}(R)$. However, one can directly calculate the mapping entropy from the simulation trajectories as well. In this section, we will briefly provide an alternative method to calculate the mapping entropy.

A. Directly sampling the values by propagating the CG simulations

From the obtained pairwise entropy function $S_{\mathrm{map}}^{(2)}$, one can enumerate the pairs during the CG simulation and calculate the pairwise entropy values. An ensemble average of the pairwise entropy values can be calculated over the propagated CG trajectories by naïvely calculating the overall mapping entropy in the same way as calculating the overall energy from pairwise contributions.

B. Calculating the ensemble averages from finite differences

Similarly, in this case, we will focus on an alternative thermodynamic quantity by

$$S_{\mathrm{map}} \approx \langle -\sum_{IJ} \frac{d\Delta W_{\mathrm{CG}}(R_{IJ})}{dT} \rangle_{T=T^*} = \left(-\frac{\Delta \langle \sum_{IJ} \Delta W_{\mathrm{CG}}(T, R_{IJ}) \rangle_{T=T^*}}{\Delta T} \right). \quad \text{(S6)}$$

Here, we note that the actual "finite difference procedure" is performed at the target temperature $T = T^*$ (= 300 K in this work) with the PMFs that are obtained at different temperatures. In other words, we calculated the finite differences of the CG free energy $\langle -\sum_{IJ} \Delta W_{\mathrm{CG}}(T, R_{IJ}) \rangle_{T=T^*}$ term itself because in this work ΔT is fixed to 25 K. In practice, we used the PMFs from $T_{\mathrm{low}} = 250$ (K) and $T_{\mathrm{high}} = 350$ (K) to perform the CG simulation at 300 K. We then calculate the averaged mapping entropy value, resulting in 19.3 Cal/mol/K.

S3-2. Convergence of entropy profiles of the bulk liquids

Using the calculation scheme from the manuscript, i.e., by calculating the finite difference of the PMFs, we depicted the pairwise entropy function from four (or three) intervals as shown in Fig. 2a. In this section, we also calculated the blockwise entropy from the finite difference of each interval and compared each to the overall averaged entropy profile. As can be seen from both methanol (Fig. S2a) and chloroform (Fig. S2b) cases, the blockwise pairwise entropy functions are in a reasonable agreement with the overall pairwise entropy using the finite differences among the entire set of temperature intervals.

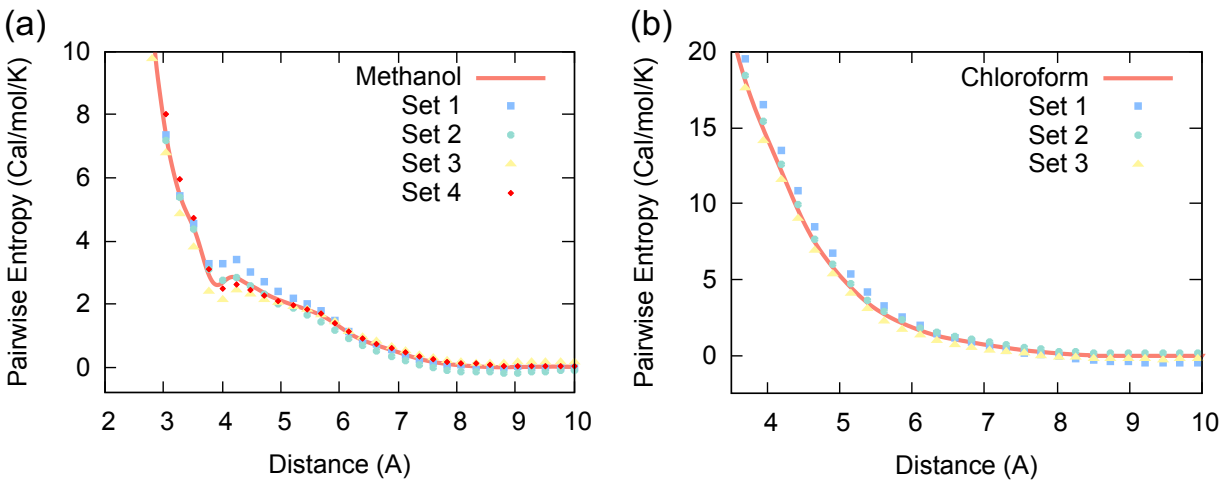

Figure S2: Convergence of pairwise entropy function $S_{\text{map}}^{(2)}(R)$ for different liquids: (a) methanol and (b) chloroform. Solid lines (red) are obtained from the finite difference of PMFs over the entire set of temperature intervals, while dots (each set represents a different temperature interval) correspond to blockwise pairwise entropy functions.

This trend is analogous to the previous report by Lu et al. using methanol.[6] However, for a system with larger deviations, we suggest utilizing the self-consistent basis (SCB) scheme[7] rather than just finite differences. As demonstrated by Wagner et al.,[7] the SCB single point formula is considered as a useful approach to calculate the sensitivity of a CG system to its underlying set of interactions. The SCB formula is followed by the reweighed force-matching residual expression with the framewise weight function defined below:

$$w_t(\mathbf{r}^n;\lambda,\delta\lambda) = \frac{\exp\left(-\beta u(\mathbf{r}^n;\lambda+\delta\lambda)+\beta u(\mathbf{r}^n;\lambda)\right.}{\frac{1}{N_t}\sum_{t=1}^{N_t}\exp\left(-\beta u(\mathbf{r}^n;\lambda+\delta\lambda)+\beta u(\mathbf{r}^n;\lambda)\right.}. \quad \text{(S7)}$$

More detailed derivation and discussion are followed in ref 7.

S4. Expression of pairwise entropy function beyond pair interactions

In this section, we present a compact description of the mapping entropy for many-body interactions. While higher-order interactions are important, we will specifically focus on three-body interactions in this section because the generalized Yvon-Born-Green (YBG) theorem[8] ensures that the MS-CG method and other bottom-up CG methods, such as g-YBG[9, 10] or iterative-YBG approaches,[11] are able to capture three-body correlations.

By including the three-body energy term, the overall energy E is now expressed as below:

$$E = \frac{1}{2}\rho^2 \int dR_{12} u^{(2)}(R_{12}) g^{(2)}(R_{12}) + E^{(3)}. \quad \text{(S8)}$$

In eq S8, $g^{(2)}$ is pair correlation function. Also, the three-body energy $E^{(3)}$ is cast as

$$E^{(3)} = \frac{\rho^3}{6V}\int dR_1 \int dR_2 \int dR_3 u^{(3)}(R_{12},R_{23},R_{31})\, g^{(3)}(R_{12},R_{23},R_{31}). \quad \text{(S9)}$$

A practical difficulty in the above expression comes from evaluating the three-body correlation function $g^{(3)}(R_{12}, R_{23}, R_{31})$. However, we can approximate $g^{(3)}(R_{12}, R_{23}, R_{31})$ as a product of pair distribution function of triplets by adopting the Kirkwood Superposition Approximation,[12] giving the integrand of $\int dR_1 \int dR_2 \int dR_3 u^{(3)}(R_{12}, R_{23}, R_{31})\, g(R_{12}) \cdot g(R_{23}) \cdot g(R_{31})$.

In practice, we can use the above formalism to calculate the "mapping entropy" of the system. Thus, the three-body energy is actually a finite difference of the PMFs over different temperatures. Based on the three-body PMF form that has previously been implemented under MS-CG theory,[13] the finite difference will have the form of

$$\Delta u^{(3)} = \Delta\left(\lambda_{IJK}\exp\left(\frac{\gamma_{IJ}\sigma_{IJ}}{R_{IJ} - a_{IJ}\sigma_{IJ}}\right)\exp\left(\frac{\gamma_{IK}\sigma_{IK}}{R_{IK} - a_{IK}\sigma_{IK}}\right) f^{(3)}(\theta_{IJK})\right). \quad \text{(S10)}$$

During the parameterization process, we can keep γ_{IK}, σ_{IK}, θ° fixed for different temperatures and only fit λ_{IJK}. Since this value is scalar (not a vector), no further complication regarding derivatives of the vector is needed.

S5. Simulation Details

S5-1. Fine-Grained Simulation (All-Atom Resolution)

A. Choice of force field

In this work, we used the Optimized Potentials for Liquid Simulations (OPLS) force field[5, 14] In the case of the neopentane and methanol systems, the AA force fields are based on ref 5. However, for the case of the chloroform system, we adopted a recent OPLS-based force field with charge corrections.[14] Implementation was aided by the use of the LigParGen framework to assign appropriate force field parameters to the system.[15]

B. Bulk liquid simulation

Initial configurations for each system were randomized after construction using the VMD[16] and Packmol[17] program packages. Atomistic force fields were chosen as described in the previous section with Ewald summation for long-range electrostatic interactions.[18] Then, we performed MD simulations with the LAMMPS simulation package.[19] We first minimized the energy to a force tolerance of 10^{-4} kcal mol^{-1} Å^{-1} with conjugant gradient minimization. Consequently, we annealed each system to its target final temperature with a Nosé-Hoover chain thermostat[20, 21] with τ_{NVT} = 0.1 ps. After reaching the desired temperature, the system was equilibrated by NPT dynamics with a Nosé-Hoover chain thermostat[20, 21] and Andersen barostat[22] at a pressure P = 1 atm and damping constant of τ_{NPT}= 1 ps. From the equilibrated system, we then performed NVT dynamics with the same thermostats and damping constants used previously. During the final NVT runs, we collected the trajectories for 5 ns with a frequency of 1 ps, which were analyzed using methods described in the main text.

C. Liquid mixture simulation

The liquid mixture was constructed by placing methanol and chloroform molecules randomly using the Packmol program.[17] An initial configuration for MD simulation was then prepared by minimizing the energy of the randomized configuration. Other simulation details remain invariant to subsection B above. We checked that the mixed phase does not phase segregate over 10 ns of simulation. This is also confirmed by examining the inter-species pair correlation (methanol-

chloroform), which has an asymptotic value of unity at large distances. More importantly, these pair correlation values match previous computational work on methanol/chloroform mixture systems.[23]

S5-2. Coarse-Grained System: Parameterization

Variational force-matching was conducted by minimizing the force differences between the FG and CG forces where the CG interactions are decomposed into spline functions

$$\chi^2[\boldsymbol{f}] = \frac{1}{3N}\langle \sum_{I=1}^{N} \left| \boldsymbol{f}_I\left(\mathbf{M}_{\mathbf{R}}^{\mathrm{N}}(\boldsymbol{r}^{\mathrm{n}})\right) - \boldsymbol{F}_I(\boldsymbol{r}^{\mathrm{n}}) \right|^2 \rangle . \tag{S11}$$

where $\boldsymbol{F}_I(\boldsymbol{r}^{\mathrm{n}})$ denotes the FG (atomistic) forces mapped onto CG bead I, while $\mathbf{f}_I\left(\mathbf{M}_{\mathbf{R}}^{\mathrm{N}}(\boldsymbol{r}^{\mathrm{n}})\right) = \boldsymbol{f}_I(\boldsymbol{R}^{\mathrm{N}}) = -\partial U_{CG}(\mathbf{R}^N)/\partial \mathbf{R}_I^{\mathrm{N}}$ denotes the effective CG force acting on the CG bead I. In this work, the CG force field utilizes a pairwise functional form defined by a set of spline functions $\{u_k\}$ and coefficients $\{c_k\}$: $\boldsymbol{f}_I(\mathbf{R}^{\mathrm{N}}) = \sum_{I\neq J}\sum_k c_k u_k(R_{IJ})\hat{e}_{IJ}$. The MS-CG method is performed for the N different FG systems with temperature $T_1, \ldots, T_N$ with a temperature spacing of $\Delta T = T_i - T_{i-1} = \tau$.

For calculating the pairwise entropy function from the CG PMF, we have actually calculated the $\Delta S_{\mathrm{CG}}(R)$ quantity using following formula:

$$\Delta S_{\mathrm{CG}}(R) = -\frac{-\int_{s=\infty}^{s=R} \frac{1}{N-1}\big(f(s, T_1 + (N-1)\tau) - f(s, T_1)\big)ds}{\tau} . \tag{S12}$$

Equation above provides a finite difference derivative of the pairwise potential, since $-\int f = V$.

S5-3. Coarse-Grained Simulation

In the current work, we utilized the multiscale coarse-graining (MS-CG) method[24-27] to parameterize the CG force field. It is worthwhile mentioning that one can also use other bottom-up CG methods such as relative entropy minimization to evaluate their fundamental methodological differences, which will be an interesting direction for future work. In the MS-CG framework, the resultant CG force field is variationally minimized by the force residuals that are shown in eq S11 of the manuscript. In practice, a linear combination of B-splines with the sixth order was fitted with a resolution of 0.20 Å to obtain the effective CG interaction. As discussed before in the manuscript, we used 20 Å as an outer cutoff of non-bonding interactions to ensure that the pairwise entropy function $\Delta S_{\mathrm{CG}}(R)$ vanishes. The inner cutoff of non-bonding interactions is dependent on the system, but we additionally fit the polynomial terms of $A \cdot \boldsymbol{R}^{-B}$ form at the inner-core region and extrapolate into hard-core region due to poor sampling.[28]

For every CG simulation, the initial configurations were obtained by mapping the last snapshot of the FG trajectories after NVT dynamics. We performed NVT dynamics with a Nose-Hoover chain thermostat[20, 21] at each desired temperature for 5 ns.

S6. Detailed Discussions on Combining Rules

S6-1. Rod-sphere interaction

As mentioned in the manuscript, we assumed that the methanol-chloroform interaction resembles rod-sphere interactions. This assumption is reasonable due to two factors. First, treating the methanol molecule as a rod-like particle is consistent with previous approaches that have used Gay-Berne ellipsoidal interactions to model methanol energetics.[29-31] Since methanol is structurally anisotropic, extending spherically symmetric single-site CG interactions to Gay-Berne interactions can better represent the methanol molecule, as shown in previous studies.[31] Second, considering the chloroform molecule as a sphere is a reasonable approximation, even though the chloroform molecule has anisotropic symmetry in contrast to the carbon tetrachloride molecule, which we did not utilize as it is immiscible with methanol. In our previous work on chloroform at the liquid/vapor interface, we demonstrated the presence of broken symmetry.[32] Nevertheless, we found out that the effect of broken symmetry (loss of isotropy) for chloroform is negligible based on the density profile observed in the liquid/vapor interface, and thus the approximation of chloroform as sphere remains valid. This argument is also supported by the energetic contribution (pair energy) of chloroform-chloroform in Fig. 2b where the peaks at local minima are not distinctly separated in contrast to that of methanol-methanol.

S6-2. Energetic combining rule

For the methanol-methanol interaction, we assumed the Gay-Berne functional form. In this document, we follow the notations and symbols from ref 31. The Gay-Berne potential between two CG sites *I* and *J* are given by following form

$$U_{GB}\left(\hat{\boldsymbol{u}}_I, \hat{\boldsymbol{u}}_J, \boldsymbol{R}_{IJ}\right) = 4\epsilon\left(\hat{\boldsymbol{u}}_I, \hat{\boldsymbol{u}}_J, \boldsymbol{R}_{IJ}\right) \times \left[\begin{array}{l} \left(\dfrac{d_w\sigma_0}{R_{IJ} - \sigma\left(\hat{\boldsymbol{u}}_I, \hat{\boldsymbol{u}}_J, \boldsymbol{R}_{IJ}\right) + d_w\sigma_0}\right)^{12} \\ -\left(\dfrac{d_w\sigma_0}{R_{IJ} - \sigma\left(\hat{\boldsymbol{u}}_I, \hat{\boldsymbol{u}}_J, \boldsymbol{R}_{IJ}\right) + d_w\sigma_0}\right)^{6} \end{array}\right], \tag{S13a}$$

$$\sigma\left(\hat{\boldsymbol{u}}_I, \hat{\boldsymbol{u}}_J, \hat{\boldsymbol{R}}_{IJ}\right) = \sigma_0\left[1 - \left\{\frac{\chi\alpha^2\left(\hat{\boldsymbol{u}}_I \cdot \hat{\boldsymbol{R}}_{IJ}\right) + \chi\alpha^{-2}\left(\hat{\boldsymbol{u}}_J \cdot \hat{\boldsymbol{R}}_{IJ}\right) - 2\chi^2\left(\hat{\boldsymbol{u}}_I \cdot \hat{\boldsymbol{R}}_{IJ}\right)\left(\hat{\boldsymbol{u}}_I \cdot \hat{\boldsymbol{u}}_J\right)\left(\hat{\boldsymbol{u}}_I \cdot \hat{\boldsymbol{u}}_J\right)}{1 - \chi^2\left(\hat{\boldsymbol{u}}_I \cdot \hat{\boldsymbol{u}}_J\right)^2}\right\}\right]^{-\frac{1}{2}}. \tag{S13b}$$

Detailed definition and functional form of each terms are detailed in ref 29, 30, and 31. In summary, the Gay-Berne potential is an anisotropic potential based on the Gaussian-overlap model where its functional form is similar to the conventional Lennard-Jones interaction but with anisotropy with not only the pair distance R_{IJ}, but with its relative orientation $\hat{\boldsymbol{R}}_{IJ}$ and each particle's orientation $\hat{\boldsymbol{u}}_I, \hat{\boldsymbol{u}}_J$ as well.

We did not explicitly fit the energetics to the general matrix equation due to its complexity. Instead, based on the energetic profile of the methanol-methanol interaction using MS-CG, i.e., $\Delta U_{\mathrm{CG}}(R)$ from eq 3 of the main text and as plotted in Fig. 2b, we obtained 1.0244 and 0.7285 kcal/mol from the first two minima of the pairwise energetics.

For chloroform, we used a modified Lennard-Jones (LJ) interaction, LJ 6-4, that was previously used to model hydrogen bonding interactions on proteins.[33] This is because the long-range interaction at larger distances (5-10 Å) does not decay rapidly, and thus the conventional attractive LJ 6 term is not able to fit this long-range behavior. By definition, the LJ 6-4 interaction has the form below, derived from the Mie Potential (n, m) with $n = 6$ and $m = 4$:

$$U_{\mathrm{CCl_3H}}^{\mathrm{LJ\ 6-4}}(R) = \frac{27}{4}\varepsilon_{\mathrm{CCl_3H}}\left(\left(\frac{\sigma_{\mathrm{CCl_3H}}}{R}\right)^6 - \left(\frac{\sigma_{\mathrm{CCl_3H}}}{R}\right)^4\right), \tag{S14}$$

giving $\varepsilon_{\mathrm{CCl_3H}} = 1.0454$ kcal/mol and $\sigma_{\mathrm{CCl_3H}} = 3.1893$ Å.

The cross-interactions between methanol and chloroform molecules were modeled using rod-sphere interactions. We specifically followed the systematic theory by Cleaver et al.[34] In the case of rod-sphere, both χ and α, which are Gay-Berne interaction parameters, go to zero (result from Table I of ref 34). Then, the resultant interaction parameters for rod-spheres are reduced to $\sigma_{r-s} = \sigma_0\left(1 - \frac{\chi}{\alpha^2}\left(\hat{\mathbf{r}}_{ij}\cdot\hat{\mathbf{u}}_j\right)^2\right)^{-\frac{1}{2}}$ and $\varepsilon_{r-s} = \varepsilon_0$ where $\frac{\chi}{\alpha^2}$ is a shape parameter. For the sake of simplicity, we further assume $\sigma_{r-s} \approx \sigma_0$. In this system, this is an acceptable approximation since $\frac{\chi}{\alpha^2} = \frac{l_r^2 - d_r^2}{l_r^2 + d^2} \approx \frac{2.55^2 - 1.00^2}{2.55^2 + (1.75\times 2)^2} = 0.29$ (distance values are measured from the atomistic structures). By adopting this approximation, the final expression for the rod-sphere interaction is given as $\sigma_{r-s} = \sigma$ and $\varepsilon_{r-s} = \sqrt{\varepsilon_r \cdot \varepsilon_s}$ for the interaction form of $U_{r-s}(R) = \frac{27}{4}\varepsilon\left(\left(\frac{\sigma_{ij}}{R}\right)^6 - \left(\frac{\sigma_{ij}}{R}\right)^4\right)$. Below is the schematic description that summarizes the energetic combining rules for the mixture system.

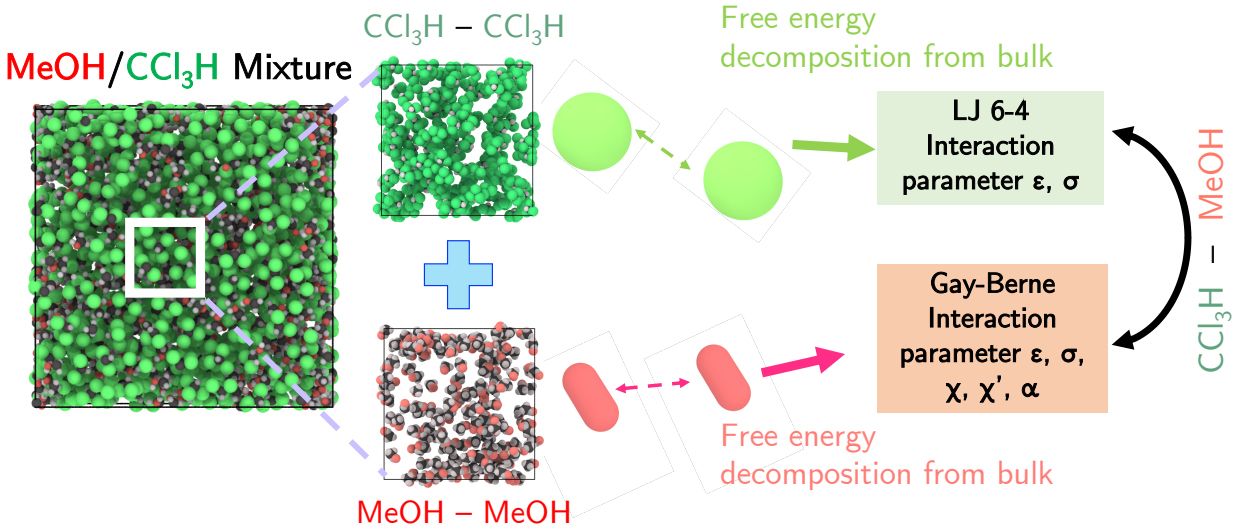

Figure S3: Schematic description of combining the energetic interactions of bulk chloroform and methanol to infer the cross-interaction in their liquid mixture. Interaction parameters are extracted from the self-interaction in bulk phases based on Lennard-Jones 6-4 interactions (chloroform) and Gay-Berne interactions (methanol), respectively.

S6-3. Entropy mixing rule

From the pairwise entropy decomposition scheme, the mapping entropy of bulk methanol (denoted as 1) and chloroform (as 2) is represented as below:

$$S_{\text{map}}^{11} = \langle \sum_{IJ} S_{11}^{(2)}(R_{IJ}) \rangle, \tag{S15a}$$

$$S_{\text{map}}^{22} = \langle \sum_{IJ} S_{22}^{(2)}(R_{IJ}) \rangle. \tag{S15b}$$

Given its near-positive definite property observed in methanol and chloroform cases (Fig. 2a in the main text), the pairwise entropy function can be written as

$$S_{11}^{(2)}(R) \approx e^{-\alpha(R)\cdot R}, \tag{S16a}$$

$$S_{22}^{(2)}(R) \approx e^{-\beta(R)\cdot R}. \tag{S16b}$$

where $\alpha(R)$ and $\beta(R)$ are functions of the distance R that can be fitted to the pairwise entropy functions from Fig. 2a. Using the interpolated functions, the entropy combining rule suggested in the manuscript can be alternatively written as below:

$$\Delta S_{\text{MeOH}-\text{CCl}_3\text{H}}^{\text{mix}}(R) = \sqrt{(\Delta S_{\text{MeOH}}^{\text{bulk}}(R)) \cdot (\Delta S_{\text{CCl}_3\text{H}}^{\text{bulk}}(R))} \approx \exp\left(-\frac{\alpha(R)+\beta(R)}{2}\cdot R\right). \tag{S17}$$

In the above equation, $S_{11}^{(2)}(R) = \Delta S_{\text{MeOH}}^{\text{bulk}}(R) > 0$ and $S_{22}^{(2)}(R) = \Delta S_{\text{CCl}_3\text{H}}^{\text{bulk}}(R) > 0$, as denoted in the manuscript. In other words, the entropy mixing rule of two pairwise entropy functions with $\alpha(R)$ and $\beta(R)$ gives the combined pairwise entropy function with $\frac{\alpha(R)+\beta(R)}{2}$ due to the near-positive definite behavior. The arithmetic mean from this combining rule is consistent with conventional Lorentz-Berthelot combining rules for σ,[35, 36] suggesting that the pairwise entropy may relate to the σ variable.

S7. CG Entropy Representability in Liquid Mixture

Similarly, we obtained the mapping entropy of the liquid mixture system by separately calculating the mapping entropy of methanol in the mixture and chloroform in the mixture. As shown in Fig. 3b of the manuscript, we also varied the cutoff value of the integration of radial entropy values and checked that $S_{\text{map}}^{\text{mix}}$ is converged, see Fig. S4 below.

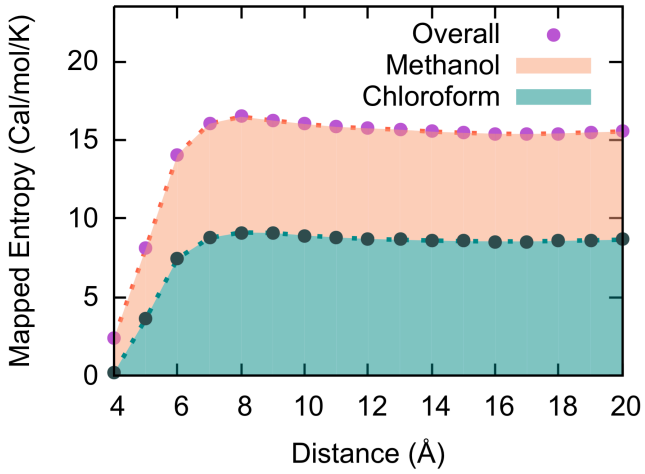

Figure S4: Mapped CG entropy obtained by integrating the radial mapped entropy $S_{\mathrm{map}}^{\mathrm{mix}}(R_{\mathrm{cut}})$ with different cutoff radii R_{cut} for the methanol/chloroform mixture system. Contributions from each liquid are marked with different colors (red for methanol and green for chloroform).

S8. Cluster Analysis Protocol

In Fig. 5c and 5d in the main text, we defined a homogeneous cluster as a fully connected graph from the network of particles within the first coordination shell. This approach was previously applied to the hydrophobic association behavior of neopentane solute in methanol solvent, clearly demonstrating the clustering behavior of neopentane.[37] Using the CG trajectory from the CG simulation, graph analyses were performed using the NetworkX package.[38]

Supplemental References

1. Wagner, J. W.; Dama, J. F.; Durumeric, A. E.; Voth, G. A., On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys.* **2016,** *145*, 044108.
2. Foley, T. T.; Shell, M. S.; Noid, W. G., The impact of resolution upon entropy and information in coarse-grained models. *J. Chem. Phys.* **2015,** *143*, 243104.
3. Koziara, K. B.; Stroet, M.; Malde, A. K.; Mark, A. E., Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *J. Comput.-Aided Mol. Des.* **2014,** *28*, 221-233.
4. Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E., An automated force field topology builder (ATB) and repository: version 1.0. *J. Chem. Theory Comput.* **2011,** *7*, 4026-4037.
5. Jorgensen, W. L.; Maxwell, D. S.; 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.
6. Lu, L.; Voth, G. A., The multiscale coarse-graining method. VII. Free energy decomposition of coarse-grained effective potentials. *J. Chem. Phys.* **2011,** *134*, 224107.
7. Wagner, J. W.; Dama, J. F.; Voth, G. A., Predicting the sensitivity of multiscale coarse-grained models to their underlying fine-grained model parameters. *J. Chem. Theory Comput.* **2015,** *11*, 3547-3560.
8. Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Voth, G. A., Multiscale coarse-graining and structural correlations: Connections to liquid-state theory. *J. Phys. Chem. B* **2007,** *111*, 4116-4127.

9. Mullinax, J.; Noid, W. G., Generalized Yvon-Born-Green theory for molecular systems. *Phys. Rev. Lett.* **2009,** *103*, 198104.
10. Mullinax, J.; Noid, W. G., A generalized-Yvon− Born− Green theory for determining coarse-grained interaction potentials. *J. Phys. Chem. C* **2009,** *114*, 5661-5674.
11. Cho, H. M.; Chu, J.-W., Inversion of radial distribution functions to pair forces by solving the Yvon–Born–Green equation iteratively. *J. Chem. Phys.* **2009,** *131*, 134107.
12. Kirkwood, J. G., Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935,** *3*, 300.
13. Larini, L.; Lu, L.; Voth, G. A., The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. *J. Chem. Phys.* **2010,** *132*, 164107.
14. Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L., 1.14* CM1A-LBCC: localized bond-charge corrected CM1A charges for condensed-phase simulations. *J. Phys. Chem. B* **2017,** *121*, 3864-3870.
15. Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L., LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017,** *45*, W331-W336.
16. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Molec. Graphics* **1996,** *14*, 33-38.
17. Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M., PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009,** *30*, 2157-2164.
18. Hockney, R. W.; Eastwood, J. W., *Computer simulation using particles*. CRC press, 1988.
19. Plimpton, S., Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995,** *117*, 1-19.
20. Nosé, S., A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984,** *81*, 511.
21. Hoover, W. G., Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **1985,** *31*, 1695.
22. Andersen, H. C., Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980,** *72*, 2384.
23. Gratias, R.; Kessler, H., Molecular dynamics study on microheterogeneity and preferential solvation in methanol/chloroform mixtures. *J. Phys. Chem. B* **1998,** *102*, 2027-2031.
24. Izvekov, S.; Voth, G. A., A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005,** *109*, 2469-2473.
25. Izvekov, S.; Voth, G. A., Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* **2005,** *123*, 134105.
26. 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.
27. 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 coarsegrained molecular models. *J. Chem. Phys.* **2008,** *128*, 244115.
28. Das, A.; Lu, L.; Andersen, H. C.; Voth, G. A., The multiscale coarse-graining method. X. Improved algorithms for constructing coarse-grained potentials for molecular systems. *J. Chem. Phys.* **2012,** *136*, 194115.

29. Berne, B. J.; Pechukas, P., Gaussian model potentials for molecular interactions. *J. Chem. Phys.* **1972,** *56*, 4213.
30. Gay, J.; Berne, B., Modification of the overlap potential to mimic a linear site–site potential. *J. Chem. Phys.* **1981,** *74*, 3316.
31. Golubkov, P. A.; Ren, P., Generalized coarse-grained model based on point multipole and Gay-Berne potentials. *J. Chem. Phys.* **2006,** *125*, 064103.
32. JinJin, J.; Voth, G. A., Ultra-coarse-grained models allow for an accurate and transferable treatment of interfacial systems. *J. Chem. Theory Comput.* **2018,** *14*, 2180-2197.
33. Fabiola, F.; Bertram, R.; Korostelev, A.; Chapman, M. S., An improved hydrogen bond potential: impact on medium resolution protein structures. *Protein Sci.* **2002,** *11*, 1415-1423.
34. Cleaver, D. J.; Care, C. M.; Allen, M. P.; Neal, M. P., Extension and generalization of the Gay-Berne potential. *Phys. Rev. E* **1996,** *54*, 559.
35. Lorentz, H., Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys.* **1881,** *248*, 127-136.
36. Berthelot, D., Sur le mélange des gaz. *C. R. Hebd. Séances Acad. Sci.* **1898,** *126*, 1703.
37. Dama, J. F.; Jin, J.; Voth, G. A., The theory of Ultra-Coarse-Graining. 3. Coarse-grained sites with rapid local equilibrium of internal states. *J. Chem. Theory Comput.* **2017,** *13*, 1010-1022.
38. Hagberg, A.; Schult, D.; Swart, P., Exploring network structure, dynamics, and function. *Proc. 7th Python Sci. Conf.* **2008**, 11-15.