Supporting Information for "Statistical Mechanical Design Principles for Coarse-Grained Interactions across Different Conformational Free Energy Surfaces"

Jaehyeok Jin† and Gregory A. Voth*
Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, IL 60637, United States
(Dated: January 31, 2023)

† Present address: Department of Chemistry, Columbia University, New York, NY 10027, United States
* Corresponding author: gavoth@uchicago.edu

I. ALL-ATOM SIMULATION DETAILS

A. Bulk Methanol Liquid

The bulk liquid methanol system was prepared by following the identical procedure described in Ref. 1. Here, we briefly summarize this procedure. The initial system with periodic boundary conditions was constructed by placing 1000 molecules in a cubic box using the Packmol program package.[2] The randomized atomistic configurations were relaxed before propagating the molecular dynamics (MD) simulation. All-atom interactions were modeled using the optimized potentials for liquid simulations all-atom (OPLS-AA) force field[3, 4] with long-range electrostatics.[5] After energy-minimization, the simulation box was annealed up to the selected temperature (200 and 400 K for obtaining the self-interactions and 300 K as our target temperature) for 0.1 ns under constant *NVT* dynamics using Nosé-Hoover thermostat,[6, 7] and then constant *NPT* dynamics were performed under 1 atm condition for 0.1 ns using Andersen barostat.[8] Finally, we ran constant *NVT* dynamics for 5 ns to obtain the trajectories for further analysis. This protocol provides CG interactions that can represent bulk liquid and gas phases at different system sizes corresponding to the constant *NPT* condition from the atomistic level. Throughout the simulation, Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) engine[9-11] was utilized with an MD timestep of 1.0 fs as well as coupling constants of 0.1 ps for the thermostat and 1.0 ps for the barostat.

B. Liquid-Vapor Interfaces

Liquid-vapor interfaces of methanol and benzene were prepared by following our previous protocol used for molecular interfaces (see Ref. 12). For clarity, Fig. S1 illustrates detailed procedures for the methanol interface. First, from the equilibrated bulk liquid (constructed according to the previous subsection) at 300 K, we placed a vacuum slab above and below the bulk system. In order to consider a finite size effect due to vacuum slab, the width of the slab for each system was chosen as approximately half the length of the cubic box: 41.3014 Å × 41.3014 Å × 81.3014 Å for methanol and 53.5971 Å × 53.5971 Å × 103.5971 Å for benzene. Next, imposed mechanical stresses and instabilities during the interface construction were removed after the energy minimization. Then, the constant *NVT* dynamics were performed for 2 ns to obtain the final trajectories using the same thermostat condition described above.

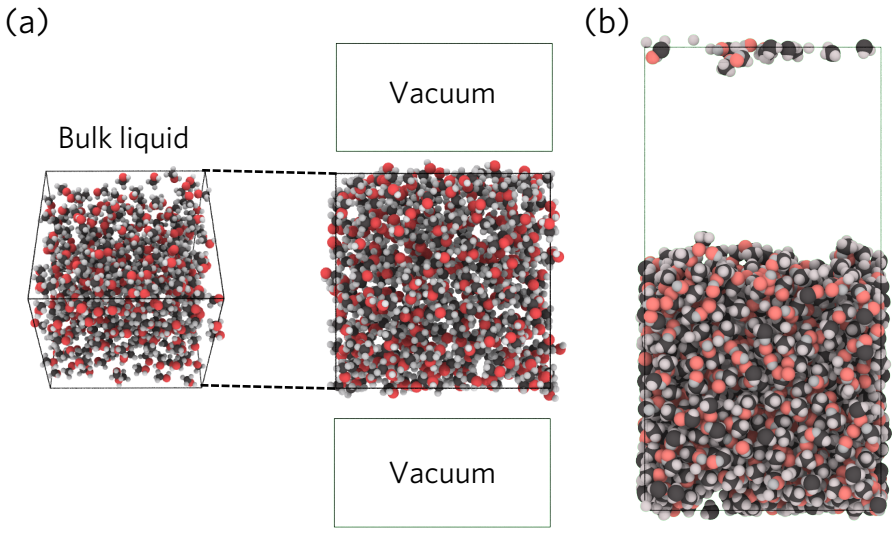

Figure S1: Construction of methanol liquid-vapor interface at the atomistic level. (a) From the equilibrated structure of bulk methanol liquid, vacuum slabs are inserted to impose the vapor interface. (b) After energy minimization, the equilibrated snapshot during the MD simulation of the interface.

C. Deca-Alanine Solvated in Water

The initial configuration for atomistic MD runs was prepared using the CHARMM-GUI[13] server, composed of one deca-alanine molecule solvated by 909 water molecules. The atomistic force field for deca-alanine was modeled using the CHARMM36 force field,[14] and the TIP3P force field[15] was used for water. Combined together, non-bonded interactions at short ranges were considered up to 12 Å. After the initial minimization process to eliminate artificial stresses, the box size was equilibrated by performing the constant *NPT* dynamics at 1 atm and 310 K for 0.5 ns using Andersen barostat.[8] The final equilibrated system size was obtained as 33.812 Å × 27.786 Å × 30.459 Å. Then, the final atomistic trajectories were sampled under constant *NVT* dynamics for 25 ns. To safely propagate the system, we used a timestep of 0.5 fs.

II. COARSE-GRAINED MODEL CONSTRUCTIONS

A. Initial Preparation of the Coarse-Graining Process

Using the produced constant *NVT* trajectories, we constructed the reference CG trajectories for parametrization by manually mapping them into the CG level. For molecular liquids, each molecule was mapped into its center-of-mass. For deca-alanine, the CG mapping only acts on the deca-alanine atoms, while we assume that the effect of waters is implicitly mapped into the deca-alanine configurations. 10-site deca-alanine CG representation was done by mapping each residue into its center-of-mass.

B. Coarse-Graining with Internal States: Bulk Liquid in Mean-Field Limit

For bulk methanol under the mean-field limit, we imposed two internal states based on our earlier report.[1] We first chose the density kernel for the semi-global number density to correctly represent the global volume fluctuations. For methanol, it is shown that a cutoff of 12 Å with a smoothly decaying density kernel $\rho_I(R_I)$ defined below can satisfy such condition:

$$\rho_I(R_I) = \sum_{IJ} \frac{1}{2}\left(1 - \tanh\left(\frac{R_{IJ} - 12}{1.2}\right)\right). \tag{S1}$$

Figure S2 depicts evaluated distributions of ρ_I at different temperatures ranging from 200 to 400 K under the constant pressure condition, which clearly shows a denser (low temperature) state derived from 200 K and a less dense (high temperature) state derived from 400 K. In this light, Ref. 1 developed a mean-field CG model with these two internal states. In order to span different temperatures, the internal state probabilities were designed based on Fig. S2:

$$p_{s_I=1,I}(R_I) = \frac{1}{2}\left(1 + \tanh\left(\frac{\rho_I(R_I) - \rho_0}{\bar{\rho}}\right)\right), \tag{S2a}$$

$$p_{s_I=2,I}(R_I) = \frac{1}{2}\left(1 - \tanh\left(\frac{\rho_I(R_I) - \rho_0}{\bar{\rho}}\right)\right), \tag{S2b}$$

where $\bar{\rho} = 10.5$ and $\rho_0 = 105$ from Ref. 1. Here, we followed the conventional definition of internal states where state *1* corresponds to a locally denser state, whereas state *2* corresponds to a locally less dense state. The internal states for the methanol liquid in this work were also designed in the same way. However, our primary focus in this paper is at 300 K condition, which is in the middle of the temperature explored in Fig. S2. Thus, we expect that our model exhibits characters of both denser and less dense states in a comparable magnitude, which is consistent with our observation illustrated in Fig. 2(b) of the main text.

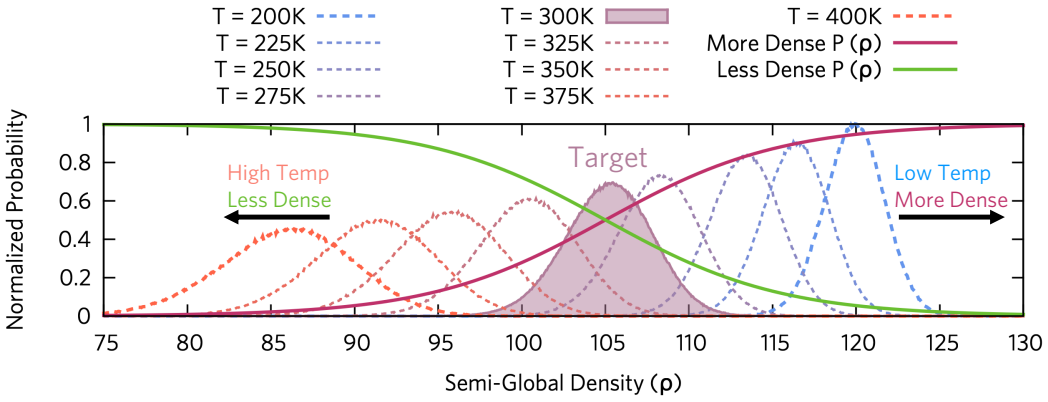

Figure S2: Determination of substate probability functional. In Ref. 1, a more dense state (magenta line) and a less dense state (green line) were determined based on the semi-global density distribution of mapped atomistic systems at temperatures ranging from 200 to 400 K. The final functional forms [Eqs. (S2a) and (S2b)] can accurately span the changes in CVs at these temperature ranges. Our target system condition of 300 K was shown in a filled curve.

C. Coarse-Graining with Internal States: Liquid-Vapor Interface

Detailed protocols of the interface CG model construction are discussed in Ref. 12 in depth. In order to correctly distinguish between the liquid-like denser state and gas-like less dense state, we utilized a local number density as a collective variable (CV). It is suggested that a cutoff distance that corresponds to the first coordination shell in liquids can be a good measure to distinguish the local environment. From the mapped atomistic radial distribution functions (RDFs) [see Fig. 3(d) of the main text], this cutoff distance R_0 corresponds to 4.5 Å for methanol and 6.0 Å for benzene. Then, the density threshold to distinguish locally denser and less dense states can be determined from the density histogram from the atomistic trajectories evaluated by

$$\rho_I(R_I) = \sum_{IJ} \frac{1}{2}\left(1 - \tanh\left(\frac{R_{IJ} - R_0}{0.1\, R_0}\right)\right). \tag{S3}$$

The computed density histograms for methanol and benzene are illustrated in Fig. 3(b) of the main text, which shows a clear separation between locally denser and less dense states. Then, the density thresholds to impose these internal states *on the fly* were chosen as $\rho_0 = 3.5$ for methanol and $\rho_0 = 4.5$ for benzene. Finally, the internal state probability functionals were implemented as

$$p_{s_I=1,I}(R_I) = \frac{1}{2}\left(1 + \tanh\left(\frac{\rho_I(R_I) - \rho_0}{\bar{\rho}}\right)\right), \tag{S4a}$$

$$p_{s_I=2,I}(R_I) = \frac{1}{2}\left(1 - \tanh\left(\frac{\rho_I(R_I) - \rho_0}{\bar{\rho}}\right)\right), \tag{S4b}$$

with a slope of $\bar{\rho} = 0.1\rho_0$ to recapitulate the gradient of density CV at the interface. Reference 12 demonstrated that the particular functional form of Eq. (S4) can capture the correct bimodal distributions between liquid-like and gas-like states, and the corresponding UCG simulation in conjunction with Eq. (3) can enhance the model performance in terms of slab density profiles and RDFs. This prior observation is consistent with the performance of the predicted CG model from the current work.

D. Coarse-Graining with Internal States: Deca-Alanine

As described in the main text, local density may not be an optimal CV for deca-alanine to correctly distinguish the folded (helix) and unfolded (coil) states. A more straightforward CV that allows for such separation is the 1-4 distances, R_{1-4}, meaning the distances between residues that take part in dihedral interactions. This argument is substantiated by the CV histogram computed for the mapped atomistic trajectories [Fig. 4(b) of the main text].

However, a single 10-site CG deca-alanine model has an overall 7 dihedral pairs with overlapping CG sites, and thus a slight modification needs to be done in order to define internal states. We circumvent this challenge by introducing internal states to only second and third residues per dihedral and permutating over all dihedrals. This design scheme allows the internal states to obey identical distributions from atomistic simulations and uniquely determine 8 CG sites for each CG configuration. Since Fig. 4(b) shows a separation between two internal states at a threshold distance of 7.0 Å, the internal state probability functionals for deca-alanine were defined as

$$p_{s_I=1,I}(R_I) = \frac{1}{2}\left(1 + \tanh\left(\frac{R_{I\in 1-4} - 7.0}{0.7}\right)\right). \tag{S5a}$$

$$p_{s_I=2,I}(R_I) = \frac{1}{2}\left(1 - \tanh\left(\frac{R_{I\in 1-4} - 7.0}{0.7}\right)\right). \tag{S5b}$$

E. General Remarks on Designing Internal States

In this subsection, we would like to provide a general remark and comment on how to design internal states and functionals for other systems for future reference. A more comprehensive review and the remark can be found in Sec. 4 of Ref. 16. Given a specific system, one first needs to define an appropriate CV that clearly shows a bimodal or multimodal trend in order to develop the corresponding multi-configurational

CG. This particular choice can be derived from the chemical or physical characteristics of the system of interest, e.g., we utilized the local (number) density and the 1-4 distance CVs in this work. This gives a definition of internal state variables $\{s_I\}$ in Eq. (3) of the main text.

Then, based on the CV histogram computed from the atomistic reference using the chosen CV, one should design the functional forms with hyperparameters of the state probability functional $p_I(s_I)$ to distinguish the multi-configurational characteristics. The imposed probability functional should have correct asymptotics, and the crossover between internal states should match the region observed from the atomistic CV histogram. Also, the derived multi-configurational CG force field form [Eq. (3)] enforces that $p_I(s_I)$ should be differentiable so that the effective force $-\nabla U_{\text{multi}}(\mathbf{R}^N)$ can be well-defined.[17, 18]

From our previous works,[1, 12, 17-19] we discovered that a hyperbolic tangent functional with the hyperparameters derived from the atomistic bimodal CV distributions can satisfy the aforementioned criteria and can correctly capture the multi-configurational nature at the CG resolution. Nevertheless, other functional forms of CVs and state probabilities should work in principle as long as the correct heterogeneous distributions in CV spaces can be captured. For example, when computing the local density CV, other functional forms with a sigmoidal profile, e.g., the Lucy function[20] widely used in smooth particle hydrodynamics,[21] have been also utilized in designing CG models.[22] Detailed analysis of the order parameter-dependent CG interactions can be found in Ref. 23.

III. PARAMETRIZATION OF COARSE-GRAINED INTERACTIONS

A. Multiscale Coarse-Graining (MS-CG) Methodology

Effective interactions between the CG particles are variationally determined by the MS-CG force-matching.[24-28] In essence, MS-CG force-matching variationally minimizes the force differences between the FG references and the unknown forces acting on the CG particle, which can be formally written in terms of force residual functional, $\chi^2[\mathbf{F}]$,

$$\chi^2[\mathbf{F}] = \frac{1}{3N}\left\langle \sum_{I=1}^{N} \left| \mathbf{f}_I(\mathbf{r}^n) - \mathbf{F}_I\left(M_{\mathbf{R}}^N(\mathbf{r}^n)\right) \right|^2 \right\rangle, \tag{S6}$$

which is a difference between the reference microscopic forces acting on the CG site *I* and the CG force field $\mathbf{F}_I\left(M_{\mathbf{R}}^N(\mathbf{r}^n)\right)$ specified by the CG mapping operator $M_{\mathbf{R}}^N: \mathbf{r}^n \rightarrow \mathbf{R}^N$. While a complete determination of the CG force field is computationally prohibited due to the many-body nature, the pairwise approximation [Eq. (2) of the main text] can reduce Eq. (S6) into a tractable form. In practice, the pairwise CG forces spanned by the pairwise basis sets $\{\phi_2(R_{IJ})\}$ are determined by utilizing B-splines $\{u_k\}$

$$\mathbf{F}_I(\mathbf{R}^N) = \sum_{J \neq I} \phi_2(R_{IJ})\hat{e}_{IJ} = \sum_k c_k \sum_{J \neq I} u_k(R_{IJ})\, \hat{e}_{IJ}. \tag{S7}$$

Combining Eq. (S7) with Eq. (S6) results in a normal equation, which is variationally solvable.[28] In practice, we used a 6th order B-splines with a resolution of 0.50 Å for molecular liquids and 0.20 Å for deca-alanine. The outer cutoff distances for pairwise interactions were chosen from both the interaction profiles and system sizes: 18 Å for the mean-field system, 10 Å for the interfaces, and 12 Å for deca-alanine. Unlike the long-range regimes, the inner-core region of CG potentials is dominated by the repulsive interactions, which results in poor sampling.[29] To reconcile this issue, we fitted the inner-core region below the first coordination shell using a polynomial repulsive interaction.

While the bulk liquid and interface systems only exhibit the non-bonded interactions between CG particles, more complex interactions between CG particles are required for the deca-alanine system due to its bonding topology. We additionally considered bonded (1-2) and angle (1-3) interactions between CG particles and determined their interactions using the force-matching algorithm. Similar to their non-bonded interactions, a 4th order B-splines with a resolution of 0.05 Å and a 2nd order B-splines with a resolution of 2.0° were used for the bonded and angle interactions, respectively. The parametrized interactions are shown in Fig. S3.

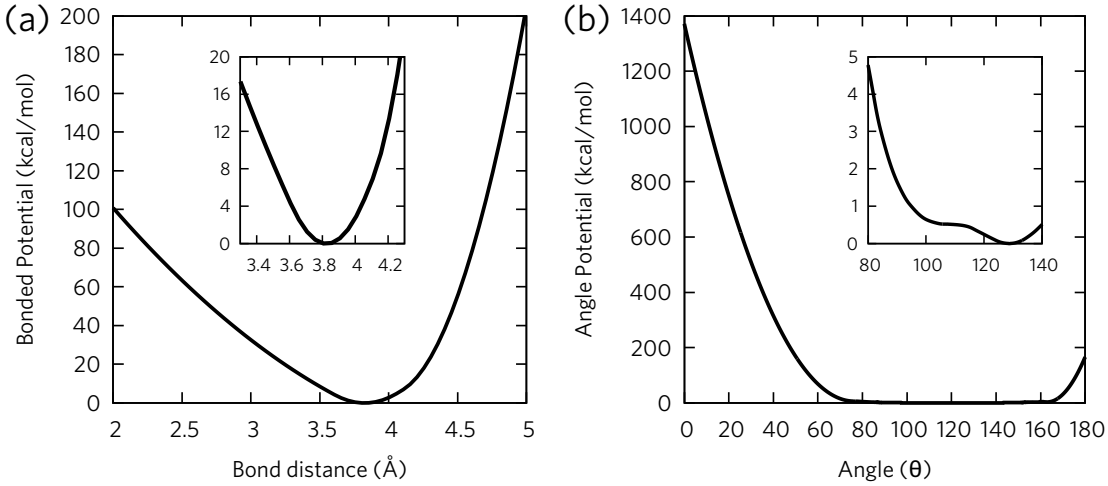

Figure S3: Parametrized CG interactions beyond pairwise non-bonded interactions for deca-alanine: (a) Bonded interaction, (b) Angle interaction.

B. Systematic Parametrization of Statewise Interactions

While Eqs. (S6)-(S7) were originally designed to uniquely determine the CG interactions between the CG particles, this variational approach can be readily extended to parametrize the statewise interactions, corresponding to distinct CG free energy surfaces divided by the chosen CV. This was proven in the earlier study,[17, 18] where the target functional is the generalized force residual:

$$\chi^2[U(\sigma,\mathbf{R}^N)] = \sum_{\sigma}\langle p_\Sigma(\sigma|\mathbf{r}^n)\|\nabla(U(\sigma,M(\mathbf{r}^n)) - M^+(\nabla U(\mathbf{r}^n) - k_B T\ln p_\Sigma(\sigma|M(\mathbf{r}^n)))\|^2\rangle_{\mathbf{r}^n}. \tag{S8}$$

In Eq. (S8), $p_\Sigma(\sigma|\mathbf{r}^n)$ denotes the state probabilities p_Σ at a given configuration $\mathbf{r}^n$, expressed as a set of state assignment $\sigma = \{s_I\}_I$, and $U\big(\sigma, M(\mathbf{r}^n)\big)$ denotes the state- and configuration-dependent CG interactions. In practice, Eq. (S8) can be interpreted as follows: the probabilistic nature of internal states can be encoded into the reference CG trajectories by constructing numerically equivalent trajectories. Hence, we implement Eq. (S8) by replicating the CG trajectories multiple times and assigning different particle types based on their state probability values. In this way, a number of CG types encountered during Eq. (S8) correctly recapitulates the internal state statistics.

IV. ESTIMATION OF THE COUPLING INTERACTION

A. Mean-Field Limit

Under the mean-field limit, we showed that solving Eq. (10) of the main text can be reduced to

$$p_1^2 U_{11}(R_{IJ}) + 2p_1 p_2 U_{12}(R_{IJ}) + p_2^2 U_{22}(R_{IJ}) = \frac{1}{2}\left[U_{11}(R_{IJ}) + U_{22}(R_{IJ})\right] - \left[\left(\frac{U_{11}(R_{IJ}) - U_{22}(R_{IJ})}{2}\right)^2 + U_{12}(R_{IJ})^2\right]^{\frac{1}{2}}. \quad \text{(S9)}$$

The additional constraint due to the UCG design principles here is $p_1 + p_2 = 1$. Therefore, in order to enforce this condition to the left-hand side, we reorganize the left-hand side in a way that has a similar term with the right hand side.

$$(LHS) = \frac{1}{2}U_{11} + \frac{1}{2}U_{22} - \left[\frac{1-2p_1^2}{2}U_{11} + \frac{1-2p_2^2}{2}U_{22} - 2p_1 p_2 U_{12}\right] \quad \text{(S10)}$$

Now, introducing the UCG constraint, it can be further simplified as

$$(LHS) = \frac{1}{2}U_{11} + \frac{1}{2}U_{22} - \left[\frac{2p_1 p_2 + p_2^2 - p_1^2}{2}U_{11} + \frac{2p_1 p_2 + p_1^2 - p_2^2}{2}U_{22} - 2p_1 p_2 U_{12}\right] \quad \text{(S11)}$$

When compared to the right-hand side of Eq. (S9), this suggests that:

$$\left[\frac{2p_1 p_2 + p_2^2 - p_1^2}{2}U_{11} + \frac{2p_1 p_2 + p_1^2 - p_2^2}{2}U_{22} - 2p_1 p_2 U_{12}\right]^2 = \left[\left(\frac{U_{11} - U_{22}}{2}\right)^2 + U_{12}^2\right] \quad \text{(S12)}$$

Using the quadratic formula, we arrive at Eq. (13) of the main text. In order to estimate $U_{12}(R)$ from Eq. (13), estimating the average of $\langle p\rangle(1-\langle p\rangle)$, $\langle p\rangle^2(1-\langle p\rangle)^2$, $\langle p\rangle^3(1-\langle p\rangle)$, and $\langle p\rangle(1-\langle p\rangle)^3$ is required. Yet, due to the mean-field nature, these values can be analytically computed from Eqs. (S2a)-(S2b) and the mean semi-global density value 105.45 from Ref. 1.

Q.E.D.

B. Beyond the Mean-Field Limit

Using the algebraic manipulations introduced in the previous subsection, solving Eq. (15) of the main text gives a solution of Eq. (16). Unlike Eq. (13), Eq. (16) contains numerous correlations between probability functionals, practically cumbersome to calculate. Given the fact that $\mathfrak{V}_1$ and $\mathfrak{V}_2$ terms have these correlations up to four-point, we instead sampled an entire range of averaged four-point correlations using the quinary variables [see Fig. S4(a)].

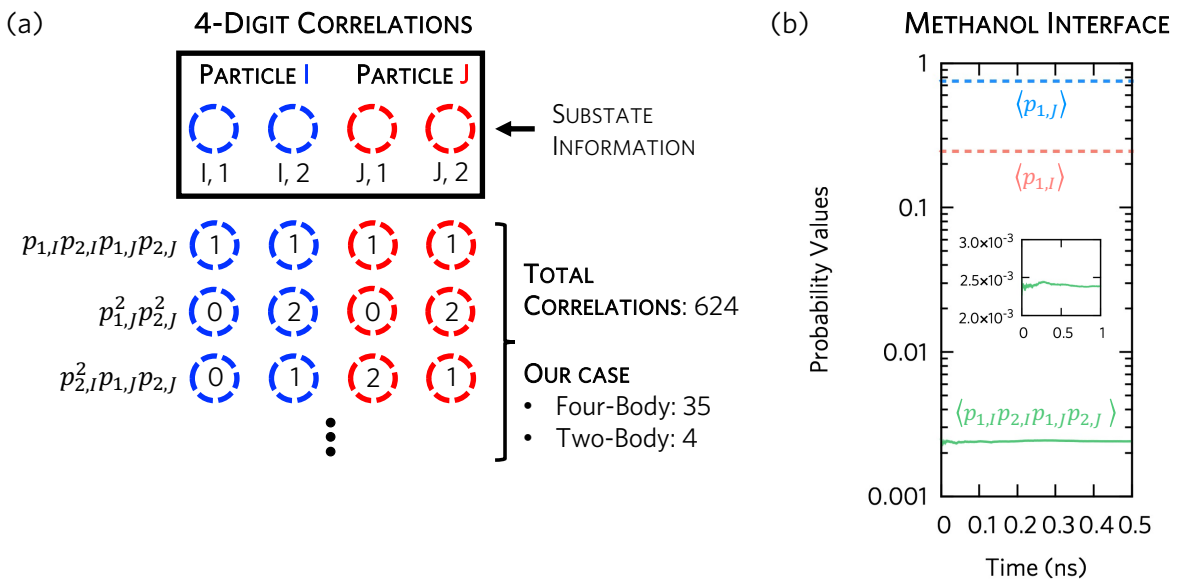

Figure S4: Implementation of efficient computation of cross-correlations between substate probabilities for CG particle pair *I, J*. (a) By introducing 4-digit quinary variables, all possible cross-correlations between internal states can be readily computed from the mapped atomistic trajectory that can be used to evaluate the cross-interaction between the CG free energy surfaces. (b) Convergence of the ensemble average of $\langle p_{1,I}p_{2,I}p_{1,J}p_{2,J}\rangle$ for the methanol liquid-vapor interface (green) described in Fig. 2. After 300 ps, the averaged correlation readily converges (see the inset). The average value of $\langle p_{1,I}\rangle = 0.246$ and $\langle p_{1,J}\rangle = 0.754$ are also plotted to demonstrate the coupled nature of probability functions.

In design, each of the four-digit quinary number corresponds to specific internal state probability correlations among CG particles *I* and *J*. The first two digits denote particle *I's* internal states (*I*, 1) and (*I*, 2), and the next two digits represent that of particle *J*: (*J*, 1) and (*J*, 2). Hence, $abcd_{(5)}$ faithfully corresponds to an ensemble-averaged value of $p_{1,I}^a p_{1,J}^b p_{2,I}^c p_{2,J}^d = \langle p_{1,I}^a p_{1,J}^b p_{2,I}^c p_{2,J}^d\rangle_{\mathrm{CG}}$. Efficient tracking of the ensemble averages for state probability correlations is needed since the state probabilities are not statistically independent in general, i.e., $\langle p_{1,I}^a p_{1,J}^b p_{2,I}^c p_{2,J}^d\rangle_{\mathrm{CG}} \neq \langle p_{1,I}^a\rangle_{\mathrm{CG}}\langle p_{1,J}^b\rangle_{\mathrm{CG}}\langle p_{2,I}^c\rangle_{\mathrm{CG}}\langle p_{2,J}^d\rangle_{\mathrm{CG}}$, unlike the mean-field case where each state probability effectively follows the Gaussian distribution. This coupled nature can be seen in Fig. S4(b), where the estimated $\langle p_{1,I}p_{2,I}p_{1,J}p_{2,J}\rangle = 0.00238$ for the methanol liquid-vapor interface is an order of magnitude lower than the independently estimated value, 0.0344.

Therefore, we implemented a scheme shown in Fig. S4(a), where a total of 624 cross-probability correlations ranging from $0001_{(5)}$ to $1111_{(5)}$ were sampled for each frame from the atomistic reference. Since $\mathfrak{V}_1$ and $\mathfrak{V}_2$ terms appeared in Eq. (16) in the main text only contain four two-body terms ($p_{1,I}p_{2,J}$, $p_{2,I}p_{1,J}$, $p_{1,I}p_{1,J}$, and $p_{2,I}p_{2,J}$) and a total of $4 + \frac{4!}{2!} + \frac{4!}{2!2!} + \frac{4!}{2!} + 1 = 35$ four-body terms, we can easily compute and filter out terms in Eqs. (17a) and (17b) of the main text. The average values converged after 300 ps simulations for both bulk liquids and interfaces (out of 2 ns and 5 ns reference) and after 5 ns for deca-alanine (out of 50 ns reference). The convergence profile of the cross-correlation function is depicted in Fig. S4(b) for $\langle p_{1,I}p_{2,I}p_{1,J}p_{2,J}\rangle$ from the methanol liquid-vapor interface as an example.

SUPPLEMENTAL REFERENCES

1. Jin, J.; Yu, A.; Voth, G. A. Temperature and phase transferable bottom-up coarse-grained models. *J. Chem. Theory Comput.* **2020**, *16* (11), 6823-6842.
2. 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* (13), 2157-2164.
3. 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* (45), 11225-11236.
4. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105* (28), 6474-6487.
5. Hockney, R. W.; Eastwood, J. W. *Computer simulation using particles*; crc Press, 1988.
6. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81* (1), 511-519.
7. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31* (3), 1695.
8. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72* (4), 2384-2393.
9. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1-19.
10. Brown, W. M.; Wang, P.; Plimpton, S. J.; Tharrington, A. N. Implementing molecular dynamics on hybrid high performance computers–short range forces. *Comput. Phys. Commun.* **2011**, *182* (4), 898-911.
11. Brown, W. M.; Kohlmeyer, A.; Plimpton, S. J.; Tharrington, A. N. Implementing molecular dynamics on hybrid high performance computers–particle–particle particle-mesh. *Comput. Phys. Commun.* **2012**, *183* (3), 449-459.
12. Jin, J.; Voth, G. A. Ultra-coarse-grained models allow for an accurate and transferable treatment of interfacial systems. *J. Chem. Theory Comput.* **2018**, *14* (4), 2180-2197.
13. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. Charmm-gui: A web-based graphical user interface for charmm. *J. Comput. Chem.* **2008**, *29* (11), 1859-1865.
14. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell Jr, A. D. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257-3273.
15. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926-935.
16. Jin, J.; Pak, A. J.; Durumeric, A. E.; Loose, T. D.; Voth, G. A. Bottom-up coarse-graining: Principles and perspectives. *J. Chem. Theory Comput.* **2022**, *18* (10), 5759-5791.
17. 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* (3), 1010-1022.
18. Dama, J. F.; Jin, J.; Voth, G. A. Correction to the theory of ultra-coarse-graining. 3. Coarse-grained sites with rapid local equilibrium of internal states. *J. Chem. Theory Comput.* **2018**, *14* (4), 2288-2288.
19. Jin, J.; Han, Y.; Voth, G. A. Ultra-coarse-grained liquid state models with implicit hydrogen bonding. *J. Chem. Theory Comput.* **2018**, *14* (12), 6159-6174.
20. Lucy, L. B. A numerical approach to the testing of the fission hypothesis. *Astron. J.* **1977**, *82*, 1013-1024.
21. Gingold, R. A.; Monaghan, J. J. Smoothed particle hydrodynamics: Theory and application to non-spherical stars. *Mon. Not. R. Astron. Soc.* **1977**, *181* (3), 375-389.
22. DeLyser, M. R.; Noid, W. G. Analysis of local density potentials. *J. Chem. Phys.* **2019**, *151* (22), 224106.

23. Wagner, J. W.; Dannenhoffer-Lafage, T.; Jin, J.; Voth, G. A. Extending the range and physical accuracy of coarse-grained models: Order parameter dependent interactions. *J. Chem. Phys.* **2017**, *147* (4), 044113.
24. Izvekov, S.; Voth, G. A. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* **2005**, *123* (13), 134105.
25. Izvekov, S.; Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109* (7), 2469-2473.
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* (24), 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 coarse-grained molecular models. *J. Chem. Phys.* **2008**, *128* (24), 244115.
28. Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. Efficient, regularized, and scalable algorithms for multiscale coarse-graining. *J. Chem. Theory Comput.* **2010**, *6* (3), 954-965.
29. 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* (19), 194115.