The Journal of Chemical Physics

RESEARCH ARTICLE | JUNE 11 2014

Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion

Timothy C. Moore; Christopher R. Iacovella; Clare McCabe

J. Chem. Phys. 140, 224104 (2014)
https://doi.org/10.1063/1.4880555

Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion

Timothy C. Moore,[1,2] Christopher R. Iacovella,[1,2,a)] and Clare McCabe[1,2,3,a)]
[1]*Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235, USA*
[2]*Vanderbilt Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, Tennessee 37235, USA*
[3]*Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, USA*

(Received 21 March 2014; accepted 19 May 2014; published online 11 June 2014)

In this work, an extension is proposed to the standard iterative Boltzmann inversion (IBI) method used to derive coarse-grained potentials. It is shown that the inclusion of target data from multiple states yields a less state-dependent potential, and is thus better suited to simulate systems over a range of thermodynamic states than the standard IBI method. The inclusion of target data from multiple states forces the algorithm to sample regions of potential phase space that match the radial distribution function at multiple state points, thus producing a derived potential that is more representative of the underlying interactions. It is shown that the algorithm is able to converge to the true potential for a system where the underlying potential is known. It is also shown that potentials derived via the proposed method better predict the behavior of n-alkane chains than those derived via the standard IBI method. Additionally, through the examination of alkane monolayers, it is shown that the relative weight given to each state in the fitting procedure can impact bulk system properties, allowing the potentials to be further tuned in order to match the properties of reference atomistic and/or experimental systems. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4880555]

I. INTRODUCTION

The utility of coarse-grained (CG) forcefields for use in simulations of soft matter and biological systems has been well established in the literature, enabling simulation to explore greater length- and time-scales than is feasible with fully atomistic models. This is of particular importance when studying the self-assembly of soft matter systems, where the assembly is typically driven by weak forces (e.g., hydrophobicity and entropy)[1–6] and structures often demonstrate hierarchical ordering (e.g., molecules organized into micelles, micelles organized into local/global patterns).[5,7–11] While generic, non-specific CG models have been widely applied,[12–16] providing important information regarding trends and design rules, it is often necessary to use CG models specifically mapped to the system of interest to provide a direct one-to-one correspondence with experiment. While several "transferable" CG forcefields, such as TraPPE-CG[17] and MARTINI,[18] have been developed, akin to forcefield development at the atomistic level,[19–23] the development of new CG forcefields is still often necessary. This is often required since the available forcefields may be lacking the necessary molecular species/groupings or may not have been optimized for the properties of interest. This second point is of particular consequence, since, for example, a forcefield optimized to match phase behavior may not appropriately capture subtle structural features.[17] Generally speaking, direct structural correspondence is needed to accurately transition between different simulation levels (e.g., atomistic to CG), in order to perform multiscale[24–26] and hybrid-multiscale simulations,[27–33] as well as to recover atomistic details from CG simulations.[34–38]

Several approaches have been developed to derive and optimize CG forcefields.[39–45] Among these, the iterative Boltzmann inversion (IBI) method[39] has become a popular choice due to its straightforward nature, general applicability to a wide range of systems, and basis in structural properties. The IBI method relies on self-consistently adjusting a given potential to achieve convergence with target structural data; for nonbonded interactions this target data takes the form of the radial distribution function (RDF) between interaction sites and the potential is iteratively updated according to

$$V_{i+1}(r) = V_i(r) - \alpha k_B T \ln\left[\frac{g^i(r)}{g^*(r)}\right], \tag{1}$$

where $V_i(r)$ is a numerical pair potential; i represents the current iteration; α is a damping factor to suppress large changes to the potential update typically varying from 0.2 to unity, where smaller values tend to be necessary to capture dense and/or crystalline states);[46] k_B is the Boltzmann constant; T is the absolute temperature; r is the separation between particles; $g^i(r)$ is the pair RDF from the simulation of $V_i(r)$, and $g^*(r)$ is the RDF of the target system mapped to the CG level. Although the CG potentials derived from IBI are typically able to accurately reproduce the target RDFs, they are, in general, only applicable at the state point for which they were derived, due to the structural nature of their derivation (e.g., note the explicit temperature dependence of Eq. (1), as well as the implicit temperature and density dependence through the $g(r)$

a)Author to whom correspondence should be addressed. Electronic mail: c.mccabe@vanderbilt.edu, christopher.r.iacovella@vanderbilt.edu

terms).[47,48] For example, separate potentials were required to capture both the solid and fluid structures of a pure simple lipid.[46] Additionally, several works have shown that CG polymer potentials derived via the IBI method can depend on the chemical environment for which they were derived.[39,50,51] Qian *et al.*[49] found that the potentials derived using IBI for ethylbenzene scale in a nonlinear fashion with temperature (i.e., a square root dependence), however, in the same work, it was found that CG potentials derived for polystyrene were independent of temperature. Other recent work has shown that some of the CG potentials in a benzene-urea-water system derived via IBI possess a degree of state point transferability, but not all.[52] It is thus unclear why IBI provides transferability for some but not all potentials; this may ultimately depend on a number of factors, including the effective size of the CG beads and the "shape" of the RDFs being fit, but may also be a limitation of the single state methodology. Furthermore, for complex systems, it may not be possible to optimize potentials at the state points of interest, due to time- or length-scale limitations of the atomistic simulations, thus making it difficult to apply the IBI method appropriately given that potentials are not necessarily transferable. Perhaps of most concern is the fact that the IBI method does not guarantee a unique solution, as a multitude of vastly differing potentials may give rise to otherwise matching RDFs. The form of the final derived potential may also vary based on runtime parameters, such as the initial potential guess, potential cutoff, magnitude of the damping factor, etc. Additionally, the derived potential may include artifacts associated with intermediate and long-range structural correlations in the system, e.g., oscillatory behavior in the potential that follows the peaks and valleys in the RDF, which may alter other properties of the system, even if the RDFs match.

In this work, the IBI method is extended to perform multi-state optimization, i.e., the potential is self-consistently adjusted to achieve simultaneous convergence of target data from multiple states. The general idea, illustrated in Figure 1, is that the inclusion of target data from multiple states adds constraints to the optimization problem, such that the derived forcefield tends toward a single potential that can adequately represent all states. For example, potentials in region *i* of the upper portion of Figure 1 are able to match the target structure at a single state *i*, potentials in region *ii* are able to reproduce target data at state *ii*, etc., with a single representative potential lying at the overlap of these regions, shown as region *iv*. To test the efficacy of the proposed multi-state iterative Boltzmann inversion (MS IBI) method, in Sec. III A, we first perform potential optimizations for the idealized system of a Lennard-Jones (LJ) fluid for which the potential is known, in order to determine if the method resolves the correct potential. In Sec. III B, to test the method in a system where only nonbonded interactions are present in the CG model, a 3-to-1 mapped CG forcefield is optimized for propane using target data generated from united-atom (UA) propane simulations, and compared with a single-site LJ model mapped to the experimental critical point of propane. In Sec. III C, we apply this approach to *n*-dodecane, a system more representative of the typical application of a CG forcefield. In Sec. III D, we examine a monolayer system composed of *n*-dodecane, where it is demonstrated that adjustment of the relative weights given to each target in the MS IBI method can be used to tune the potentials to match other measurable system properties beyond the RDF.

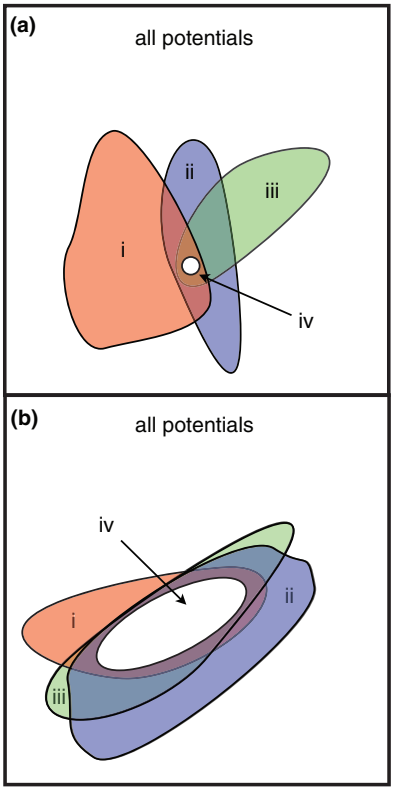

FIG. 1. Regions of good potential phase space for states with optimal overlap (top) and too much overlap (bottom).

II. METHOD AND SIMULATION DETAILS

A. Single-state iterative Boltzmann inversion

In the IBI method (which for clarity we shall refer to as single state, SS IBI), a numerical pair potential, $V(r)$ is iteratively updated according to Eq. (1). In this manner, $V(r)$ is updated at each separation, r, based on whether the RDF overpredicts or underpredicts the target RDF at the given r, and is repeated until the trial RDF matches the target RDF within some tolerance.[39] The initial guess of the numerical potential is often taken to be the Boltzmann inversion of the RDF of the target system

$$V_0(r) = -k_B T \ln(g^*(r)). \quad (2)$$

While not exact for site-site interactions in molecules,[53] this methodology is motivated by the statistical mechanics relationship between the potential of mean force (PMF) and the RDF, and provides a reasonable starting potential over which to iterate.

Typically, potentials derived with this method are capable of reproducing the target RDFs with high accuracy, with slight deviations resulting from information lost during coarse-graining. The ease of use of the IBI method and its general applicability make it a powerful tool; given a CG

mapping and a target RDF, site-site pair potentials can be readily derived with little user input.

B. Multistate extension of IBI

Although potentials derived with SS IBI will typically reproduce their target RDFs with high accuracy, caution must be taken when using the potentials. Upon successful convergence of the potential, it is only guaranteed that the derived and target RDFs match, not that the potential is necessarily representative of the "true" underlying potential (i.e., not necessarily state independent). It is important to note, especially since information is lost due to coarse-graining, that a multitude of potentials may give rise to similar RDFs. Only a small portion of the potentials that produce matching RDFs may actually fall within the region of potentials that match the true potential and, since the true potential is typically unknown, it is difficult to assess the accuracy of the derived potentials. If the derived potential falls far outside the true potential region, this may give rise to potentials that, despite providing a good match for the target RDF, lack transferability and may contain artifacts making them incapable of resolving system properties other than the RDF.

The proposed MS IBI method aims to minimize the state dependence of the derived potentials by adding additional constraints to the optimization process such that the derived potentials fall within the region of phase space where potentials are representative of the "true" potential. This approach relies on two key assumptions: (1) different thermodynamic states have different regions of the potential phase space that adequately reproduce their respective target RDFs, and (2) that the true, underlying potential lies within the common overlap between these regions of phase space. As the name suggests, this is accomplished by updating the derived potential to simultaneously match target RDFs at different thermodynamic states, producing a single potential that provides sufficient matching for all target RDFs considered. As shown graphically in the upper portion of Figure 1, the converged potential lies at the intersection of each of the regions representing the target RDFs, as this is the only region where a sufficient match will be found for all states.

The implementation of MS IBI is similar to that of SS IBI, the only additional requirement is more target data. As in SS IBI, the initial potential is assumed to be the Boltzmann inversions of the target RDFs, averaged over the N states used,

$$V_{s,0}(r) = -\frac{1}{N}\sum_{s} k_B T_s \ln(g_s^*(r)), \tag{3}$$

where the subscript s represents the property as state s. After a trial CG simulation is run at each state using the potential from Eq. (3), the potential is updated according to

$$V_{s,i+1}(r) = V_{s,i}(r) - \frac{1}{N}\sum_{s} \alpha_s(r)\, k_B T_s \ln\left[\frac{g_s^i(r)}{g_s^*(r)}\right]. \tag{4}$$

While in SS IBI, α represents a damping factor useful for suppressing fluctuations in the potential update, here $\alpha_s(r)$ also serves as a weighting factor, allowing more or less emphasis to be put on each state. For example, if fitting a potential with three states, where state 1 will ultimately be of most interest, it may make sense to give state 1 a higher α value; this will be discussed later in Sec. III D. Additionally, here $\alpha_s(r)$ is defined as a linear function of the separation r, with the points $\alpha_s(0) = \alpha_{max}$, and $\alpha_s(r_{cutoff}) = 0$. When $V_s(r_{cutoff}) = 0$ is used for the initial potential guess, this ensures that the derived potential smoothly decays to zero at the interaction cutoff, r_{cutoff} (i.e., the point at which we assume that pair interactions are zero). Since α decreases as r increases, increased emphasis is placed on shorter-range interactions compared to long-range interactions, similar to the radial dependence of the pressure correction formula often used with IBI.[39] This helps to suppress the influence of long-range structural correlations on the derived potential, as short-range interactions may certainly give rise to long-range correlations (e.g., the formation of bulk crystals from particles interacting through a short-ranged, truncated potential). For direct comparability in this work, both SS and MS IBI treat the damping factor as a linear function of separation, with a fixed value of 0 at the potential cutoff. Note that, although bonded interactions may be optimized in a similar manner (i.e., adjusting the potential to match a target distribution), in this work, we make the assumption that bonded and non-bonded interactions are sufficiently independent such that we use analytical bonded potentials, as has been done in previous work.[46,54,55]

The choice of states used in the fitting procedure is naturally important to deriving an accurate potential. To derive a potential most representative of the underlying one, it would not be beneficial to choose states with RDFs that are too similar, as the overlap region would be large, essentially providing minimal additional constraints; this situation is shown in the lower portion of Figure 1. In such a case, there would be no advantage to the multistate fitting. At the other end of the spectrum, there may in fact be no overlap between states, or more specifically, no overlap for a given level of matching (i.e., no overlap without relaxing the tolerance of a RDF similarity test). For some systems, it may not be possible to define a single pair potential that accurately reproduces the target structure at all states. This is not a problem unique to CG potentials, as it applies at all levels of modeling, e.g., classical atomistic potentials may also lack full state-independence given that they do not allow variation in electron density.

C. Simulation model

In this work, simulations were performed using 3 distinct models: generic LJ fluid, TraPPE-UA, and CG models derived via IBI. First, simulations of monatomic LJ spheres were performed in the canonical ensemble (i.e., fixed number of particles N, volume V, and temperature T), with temperature controlled via the Nosé-Hoover thermostat. These monatomic LJ systems contained 1468 particles initially randomly distributed throughout the box, and were run for 1×10^6 timesteps, during which the reduced temperature was decreased from 2.0 to the final target temperature. The systems were further equilibrated for 1×10^6 timesteps before target data were collected over 1×10^5 steps. A timestep of

1×10^{-3} in reduced time units was used. The interaction parameters used in all LJ simulations were $\sigma = 1.0$ and $\varepsilon = 1.0$, with a potential cutoff $r_{cutoff} = 3\sigma$. Here, no coarse-graining was applied to the target systems, as these simulations were used simply to test the efficacy of the potential derivation under the ideal circumstances where the true potential is known.

The second model used relies on the TraPPE-UA forcefield for simulation.[19] Here, alkanes were simulated in the canonical ensemble, with temperature controlled via the Nosé-Hoover thermostat. Bulk fluid systems of both propane (1024 molecules) and *n*-dodecane (400 molecules) were simulated at 3 different states, as listed in Sec. III, and used to generate target RDF data. Although not an all-atom model (as hydrogens are not explicitly modeled), the TraPPE forcefield was chosen for computational convenience, since, in principle, the target data can come from any source. In all cases, a timestep of 1 fs was used. After an initial equilibration period of 5 ns, data were collected over a 10 ns production run. In addition to the bulk fluid *n*-dodecane simulations, UA simulations were performed of *n*-dodecane gel and fluid monolayers, composed of 100 *n*-dodecane chains each. These were performed in the same manner as the bulk simulations at 298 K, but with the first bead of each chain held stationary such that a 2D hexagonally arranged periodic array with density 4.10 chains per nm^2 (gel-like) and 3.79 chains per nm^2 (fluid-like) was achieved; these were chosen to match state points commonly used in alkylsilane monolayer simulations and experiments.[56]

The third model used is a CG representation of alkanes. In all cases, a 3-to-1 CG model (i.e., each CG bead represents 3 UA carbon groups) was used to simulate bulk fluid and monolayer systems of alkanes. Pair potentials were derived using the SS and MS IBI methods, using the results of the UA simulations as target data, as discussed in detail in Sec. III. The bond stretching and angle bending potentials used in the study of dodecane were derived by a Boltzmann inversion of the bonded distributions sampled in the atomistic trajectory mapped to the CG level.[54] Specifically, from a normalized bond length distribution $p(r)$, the bond stretching potential is written as

$$V_{bond} = -k_B T \ln\left(p\left(r\right)\right), \tag{5}$$

which, assuming a Gaussian bond length distribution, results in a harmonic potential about the most probable bond length, r_{eq}; note an identical formalism was used for angles, where θ is substituted for r, and the normalization includes a factor of $\sin^{-1}\theta$. Since minimal state dependence was found between systems, a single set of bonded parameters was used in all simulations, with $k/k_B = 15.60$ K/Å and $r_{eq} = 3.56$ Å for bonds and $k/k_B = 0.17$ K/deg^2 and $\theta_{eq} = 174.53$ Å for angles. Bond histograms and additional details are included in the supplementary material.[57]

In all cases, the GPU-enabled HOOMD-Blue[58,59] simulation engine was used to perform the simulations. The high performance of the GPU allows for rapid derivation of potentials. A standard potential optimization using MS IBI required approximately 50 iterations to be well-converged. For the pure LJ systems with 1468 particles, this convergence took less than 1 h using three NVidia GTX580 GPUs concurrently. The following convergence criteria were used to measure how well a trial RDF matched with its target, where *dr* is the size of a RDF bin,

$$f_{fit} = 1 - \frac{\sum_0^{r_{cut}} \left(\left|g^i(r) - g^*(r)\right|\right)}{\sum_0^{r_{cut}} \left(\left|g^i(r)\right| + \left|g^*(r)\right|\right)}. \tag{6}$$

An f_{fit} value of unity represents a perfect match between the trial and target RDFs. Additionally, in all figures, the following two-point central moving average smoothing function was applied to the derived potential to reduce the noise

$$V_n^{'}(r) = \frac{1}{3}[V_{n-1}(r) + V_n(r) + V_{n+1}(r)], \tag{7}$$

where $V_n(r)$ is the *n*th element of the numerical potential, and the prime denotes the smoothed value. The application of the smoothing function was not found to significantly influence the behavior or degree of matching.

III. RESULTS

A. Monatomic Lennard-Jones fluid

To test the efficacy of the MS IBI method, potentials were derived using RDFs from monatomic LJ spheres as target data, and the results compared to single state potential derivation (i.e., SS IBI). Target data were acquired from the following states: state A, reduced density $\rho^* = N\sigma^3/V = 0.85$, reduced temperature $T^* = k_B T/\varepsilon = 0.5$; state B, $\rho^* = 0.67$, $T^* = 1.5$; and state C, $\rho^* = 0.18$, $T^* = 2.0$. No coarse-graining was performed since the goal was to test whether the MS IBI method could recover a known potential. In contrast to mapping an atomistic system to the CG level, no information about the system is lost ensuring that a single potential is applicable to all states and that this potential is known.

While the RDFs match well, as illustrated in Figure 2, the potentials derived via SS IBI demonstrate significant state dependence, as shown in Figure 2(d). For the more dense states A and B (Figures 2(a) and 2(b)), the SS IBI method was not able to converge to the true potential to the extent that in the most dense system (state A), the converged potential is almost purely repulsive. This result is due to the elevated density of this state, where the structure can be reproduced with a purely repulsive potential.[60] In this case, even though the RDF matches the target well, the overall behavior of the system would be dramatically altered as compared to the target. A similar situation arises in state B where only a weak attractive potential is required to match the target structure. In state C, however, the low density causes attractive forces to become important, and as such, the attractive portion of the LJ potential is needed to fully reproduce the target data. Thus, the true LJ potential is recovered only for state C. The application of SS IBI to the monatomic LJ system illustrates two points: (1) that potentials derived via SS IBI are state-dependent, and (2) these potentials are not unique, in that both the LJ potential and the vastly differing derived potential produce matching RDFs.

MS IBI aims to address the aforementioned issues by forcing the potential to sample portions of potential phase

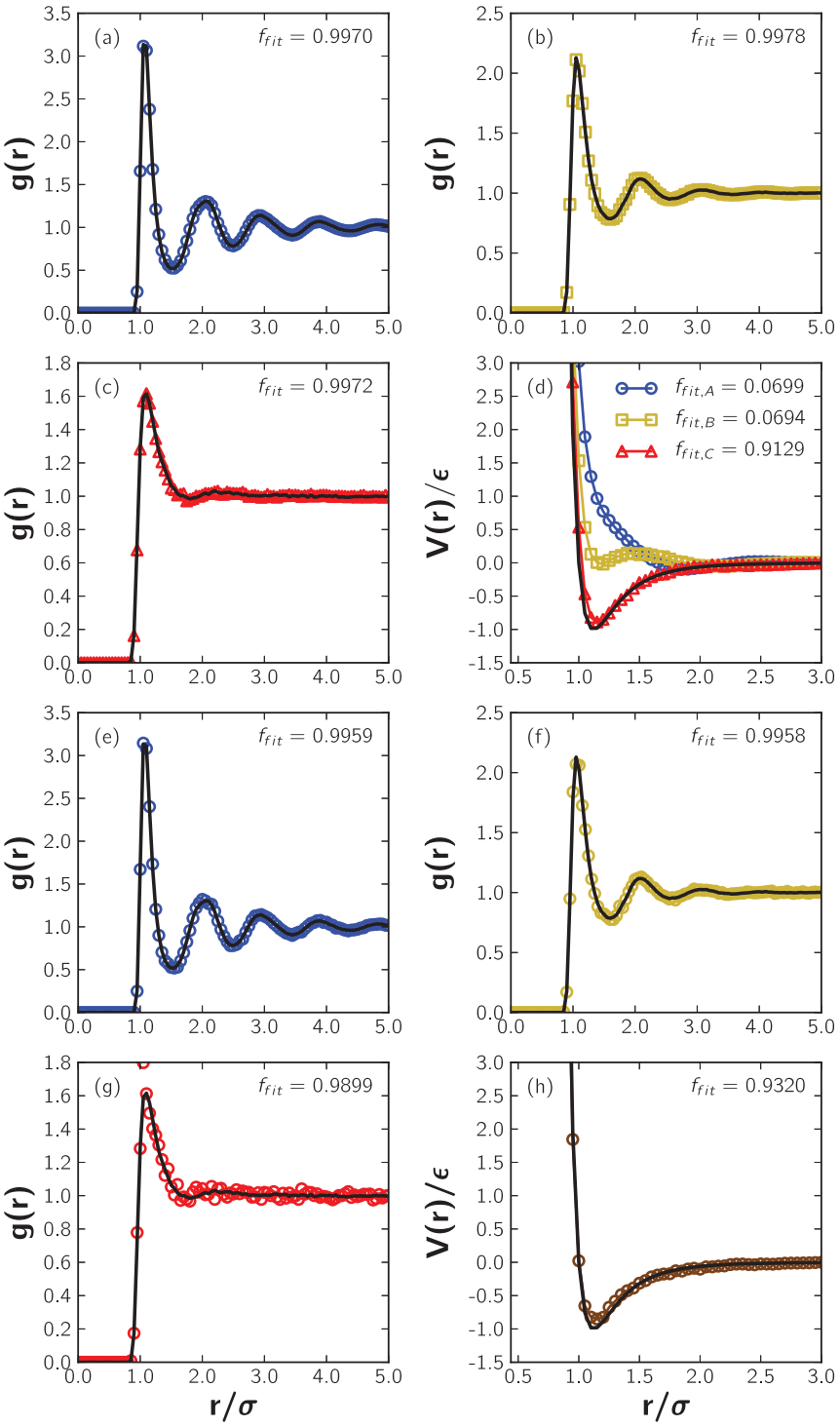

FIG. 2. RDFs and potentials derived for the LJ system. (a)–(d) SS IBI results. (e)–(h) MS IBI results. The α value used for the MS IBI optimization was 0.7 for each state. f_{fit} for the potentials was calculated in the range $\sigma \leq r \leq r_{cutoff}$. The solid black line represents the target RDF (a)–(c), (e)–(g) or the known potential (d) and (h). The circles represent the derived potential (d, h) or the RDFs calculated from simulations using the derived potential (a)–(c), (e)–(g).

space that satisfy all of the constraints, i.e., find a single potential that matches the target structure at multiple states. The results of applying MS IBI to the monatomic LJ fluid are shown in Figures 2(e)–2(h). The inclusion of target data from multiple states results in closely matching RDFs and a derived potential that accurately reproduces the true LJ potential, as shown in Figure 2(h). Although this example is simple, as no coarse-graining was performed, it illustrates the ability of MS IBI to recover a known potential and reduce state-dependence of the derived potential.

B. Propane

To further test the MS IBI algorithm, potential optimizations were performed on propane. The chosen 3-to-1 mapping results in a single-site model that can be directly compared to

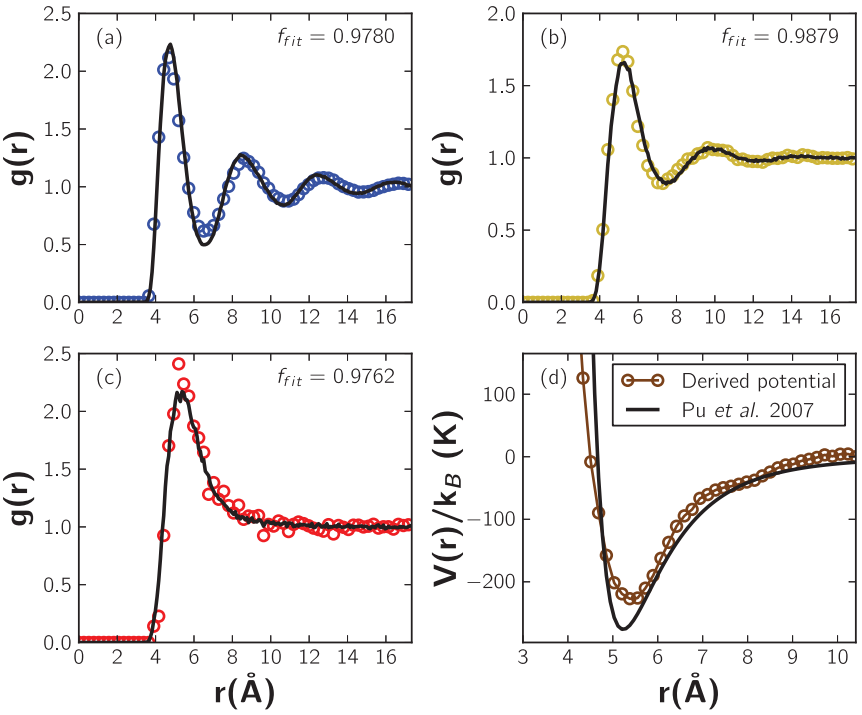

FIG. 3. RDFs (a)–(c) and potential (d) derived for propane using MS IBI. (a)–(c) correspond to states A, B, and C in the text, respectively. The α values used were 0.5, 0.7, and 0.5 for states A, B, and C, respectively.

known single site 12-6 LJ models from the literature.[61] Note, the 12-6 LJ potential should not be considered to be the "true" potential, but rather a good approximation. Target data were acquired from UA simulations at the following states: state A, 298 K, 0.818 g/ml; state B, 298 K, 0.439 g/ml; and state C, 298 K, 0.014 g/ml. The damping values used in the subsequent MS IBI optimizations were $\alpha_A(0) = 0.5$, $\alpha_B(0) = 0.7$, and $\alpha_C(0) = 0.5$. The resulting RDFs and (single) pair potential are presented in Figure 3. At each state, f_{fit} indicates excellent agreement between the target RDFs and those calculated from simulations using the derived potential. Moreover, we find that the derived potential agrees well with a single-site 12-6 LJ model using parameters mapped to the critical point of propane,[61] providing confidence in the MS IBI method. While the match between the two potentials is good, the derived potential does show two small bumps at ~7 Å and another at ~9.5 Å, which are likely related to orientational effects; that is, this is likely related to treating the three UA carbon groups as a single, spherically symmetric interaction site.

To illustrate the consistency of the potentials derived via MS IBI (i.e., that the final potential is insensitive to the initial guess), optimizations were performed using a number of different initial potentials. In addition to the PMF-like quantity of Eq. (3), three additional initial guesses were used, each a 12-6 LJ potential with vastly differing parameters: (1) $\varepsilon_1 = 0.46$ kcal/mol, $\sigma_1 = 4.51$ Å; (2) $\varepsilon_2 = 0.001\varepsilon_1$, $\sigma_2 = \sigma_1$; and (3) $\varepsilon_3 = 2\varepsilon_1$, $\sigma_3 = \sigma_1$. The final derived potentials are, in each case, very similar to each other and to the derived potential in Figure 3(d), as shown in Figure 4. Particularly, the f_{fit} values between each potential and the derived potential shown in Figure 3(d) are 0.986, 0.980, and 0.986, respectively.

C. *n*-dodecane

To examine a more complex system and test the state-independence, *n*-dodecane was examined in the bulk state. Intermolecular pair potentials were derived for the beads of a CG model of *n*-dodecane, again using a 3-to-1 mapping. The resulting 4-site model contains two middle beads and two terminal beads, where middle and terminal beads were treated as unique entities, resulting in the need to derive three pair potentials; harmonic bonds, and angles were used, as detailed in Sec. II C. The target data were collected from UA simulations of *n*-dodecane at the following states: state A, 298 K, 1.04 g/ml; state B, 298 K, 0.74 g/ml; and state C, 370 K, 0.55 g/ml; the damping values used were $\alpha_A(0) = 0.5$, $\alpha_B(0) = 0.7$, and $\alpha_C(0) = 0.5$. Note that state B corresponds to the experimental density at standard ambient temperature and pressure, and, as such, was given higher weight than the other states in this example. Close agreement with the target RDFs was found, with an f_{fit} value greater than 0.98 for each of the nine RDFs calculated (shown in the supplementary material[57]).

As a further test to assess the quality of the potentials derived via MS IBI, the average squared radius of gyration normalized by the average end-to-end distance, denoted by R_{chain}, was calculated, providing a measure of the chain conformations at different thermodynamic states. Using potentials derived with MS IBI, good agreement was seen between the UA target data and the CG model for the ratio R_{chain}, as shown in Figure 5; in this plotting scheme an ideal match corresponds to a data point situated on the line $y = x$. MS IBI deviates slightly from the target data for larger values of R_{chain}, which is likely related to the fact that, for simplicity, only a single Gaussian was used to fit the bond-angle distribution;

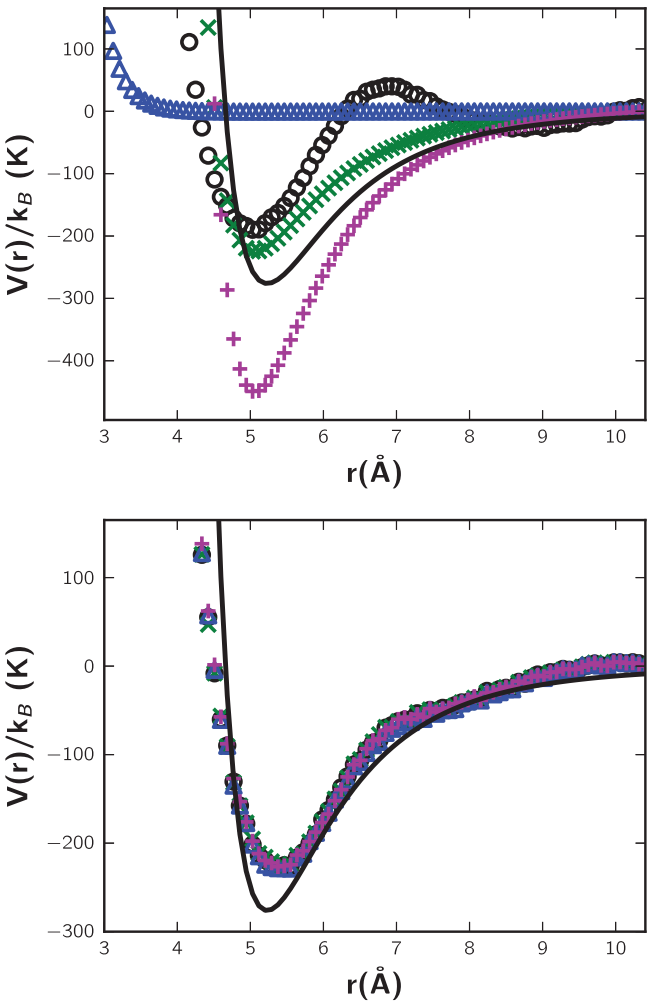

FIG. 4. Different initial guesses (top) and the resulting derived potentials (bottom) for propane optimizations. Blue triangles: LJ with $\varepsilon = \varepsilon_2$ and $\sigma = \sigma_2$; black circles: $V_0(r)$ from Eq. (2) (same shown in Figure 3); green "x": LJ with $\varepsilon = \varepsilon_1$ and $\sigma = \sigma_1$; solid black line: 1-site propane model[49] (not used as initial guess, shown for reference); magenta "+": LJ with $\varepsilon = \varepsilon_3$ and $\sigma = \sigma_3$. Symbols in top plot correspond to the same symbols in the bottom plot. Note that all potentials converge to very similar values.

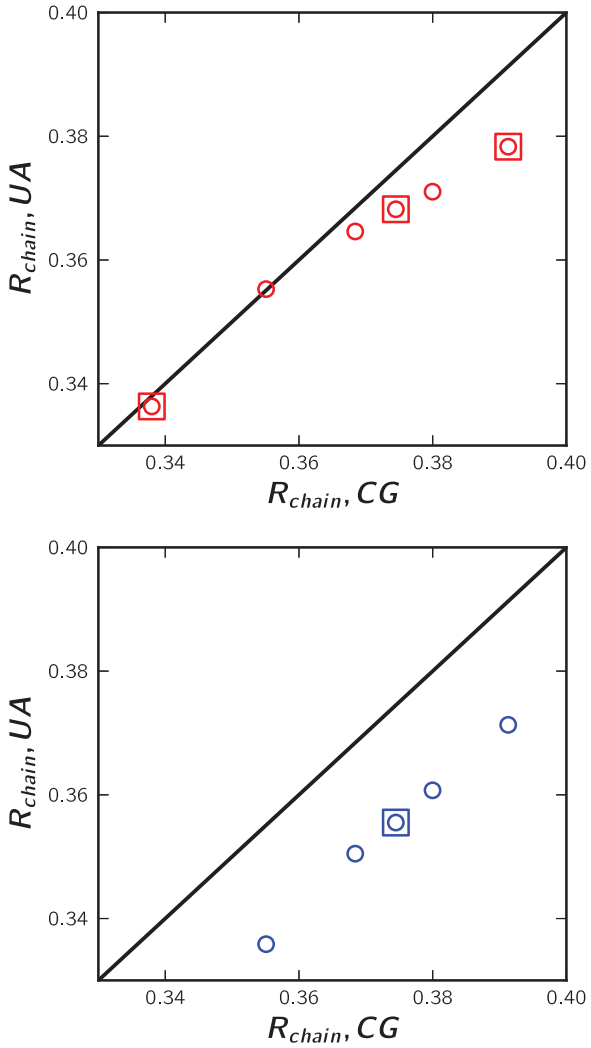

FIG. 5. Comparison of a structural metric between the CG and UA models of n-dodecane. The CG potentials were derived from MS IBI (top) and SS IBI (bottom). A value lying on the solid line represents a perfect match between the CG and UA models. Squares represent data points from simulations at state points where the potential was derived; circles are data points from other states used for testing the state dependence. The states used in the multi-state fitting are states A, B, and C as described above the n-dodecane with $\alpha_A(0) = 0.5$, $\alpha_B(0) = 0.7$, and $\alpha_C(0) = 0.5$. State B was used for the single state fitting with $\alpha = 0.7$.

for *n*-dodecane, the bond angle distributions of states A and B are closely approximated by a single Gaussian, however state C, which corresponds to larger values of R_{chain}, may be better approximated by multiple super-imposed Gaussians, as has been proposed by Milano *et al.*;[54] the use of two Gaussians would allow chains in the less dense states to adopt smaller, less extended configurations, leading to better agreement with the UA target data (bond angle histograms are included in the supplementary material[57]). While deviations for MS IBI become more apparent as R_{chain} increases, the potentials derived from only state B (i.e., standard temperature and pressure) via SS IBI show larger, systematic deviations of R_{chain} over the entire range of state points considered, in all cases, predicting larger values of R_{chain}. Since both MS IBI and SS IBI employ the same bond/angle parameters, it would have been expected that SS IBI would also show the same increased deviations for larger R_{chain}. However, the ability of MS IBI to more appropriately capture the chain conformations of *n*-dodecane appears to be directly attributable to the increased accuracy of the derived pair potentials. Note, in both cases, additional simulations were performed at state points not used in the fitting (state points used in the fittings are highlighted with open squares in Figure 5), showing good transferability of the derived potentials of MS IBI.

D. Tuning CG potentials

While it is not uncommon to tune potentials to reproduce desired behaviors, such tuning is often accomplished by manually adjusting the potentials or mixing potentials derived at different state points, e.g., modifying cross-interactions to ensure stability of CG bilayers.[62] MS IBI provides a more systematic way to optimize the CG forcefields by adjusting the relative weights applied to each state in the optimization process, providing a similar effect to manually mixing parameters from optimizations performed independently at

different state points. As mentioned in Sec. II B, here the damping coefficient, $\alpha_s(r)$, is a function of both separation, r, and state, s. Adjusting the $\alpha_s(0)$ value given to each state, s, alters the weight given to the state in the fitting, i.e., more or less emphasis can be placed on matching a given state. While adjusting the relative weights may have only a small effect on the derived RDFs, it may alter subtle features of the derived potential which can significantly alter other properties. To demonstrate this, as well as to further test the transferability of the derived potentials, alkane monolayers were simulated using the 3-to-1 CG model, with potentials optimized in the bulk states discussed above, using various values of $\alpha_s(0)$ for each of the three states. The average tilt angle, θ, with respect to the surface and the nematic order parameter, S_2,[63] of the chains were calculated and compared with those values calculated from the corresponding UA simulations. The nematic order parmeter is defined as $S_2 = \langle \frac{3}{2}\cos^2\theta - \frac{1}{2} \rangle$, where θ is the angle between an individual chain in the system and the average direction of all the chains in the system; this yields a value of unity for perfect orientational ordering, decaying to zero as ordering is lost. Note that the UA monolayer simulations were not used as target data in the potential derivation; they were used only to validate the properties predicted by the derived CG potential.

Unique sets of CG potentials were derived over a range of $\alpha_s(0)$ values, as summarized in Table I. Here, the states A, B, and C are the same states previously used as target data to derive a CG potential for bulk systems of *n*-dodecane above in Sec. III C. Initially, potentials were optimized with equal weights assigned to each state. As shown in Table I, this α set yields potentials that significantly overpredict the fluid phase order parameter, while underpredicting the gel phase chain tilt. Since the monolayers are inherently ordered, it would be expected that increasing the relative weight given to the most dense state—state A—would yield a potential that better captures the system behavior. By systematically reducing the weight given to the less dense states, first state C, then state B, a potential that very closely reproduces the monolayer behavior is obtained for $\alpha_s(0)$ values of 0.7, 0.1, and 0.1 for states A, B, and C, respectively. Given the small weights assigned to states B and C, it may be expected that this potential would give results similar to the potential derived via SS IBI at state A. However, it can be seen in Table I that this clearly is not the case; potentials derived from SS IBI at state A show large deviations, underpredicting both the average tilt angle and nematic order parameter in the gel phase monolayer, in stark contrast to the near perfect behavior predicted by MS IBI. This result is a direct consequence of using the MS IBI method; even though low weights are given to the other states, the derived potentials will only be considered converged if all states demonstrate good agreement. That is, state points with lower weights are still able to make changes to the potential, albeit of smaller magnitude; if those changes negatively impact the fitting of higher weighted states, a strong correction will be applied, producing subtle changes to the potentials as compared to systems with equal weight. The weights can also be defined as a function of the potentials for each state, e.g., equal weights for all potentials and states except the potential between terminal and middle beads for state A.

TABLE I. Average chain tilt with respect to surface normal, θ, and nematic order parameter, S2, for the monolayers in the fluid state (subscript F) and in the gel state (subscript G). The states A, B, and C are the same ones used in Subsection III C. Values are given as ensemble averages ± standard deviation.

$\alpha_A/\alpha_B/\alpha_C$	θ_F	$S_{2,F}$	θ_G	$S_{2,G}$
0.7/0.7/0.7	19 ± 6.3°	0.97 ± 0.013	18 ± 6.5°	0.98 ± 0.017
0.7/0.7/0.6	20 ± 6.1°	0.97 ± 0.012	20 ± 3.7°	0.992 ± 0.0018
0.7/0.7/0.5	18 ± 6.9°	0.96 ± 0.020	20 ± 3.8°	0.991 ± 0.0022
0.7/0.7/0.4	15 ± 7.0°	0.95 ± 0.018	20 ± 4.0°	0.990 ± 0.0027
0.7/0.7/0.3	13 ± 6.8°	0.94 ± 0.019	19 ± 4.2°	0.990 ± 0.0036
0.7/0.7/0.2	12 ± 6.3°	0.93 ± 0.014	12 ± 6.4°	0.96 ± 0.014
0.7/0.7/0.1	10 ± 5.6°	0.95 ± 0.011	8 ± 5.3°	0.958 ± 0.008
0.7/0.4/0.1	18 ± 9.0°	0.83 ± 0.021	29 ± 3.9°	0.988 ± 0.0026
0.7/0.1/0.1	20 ± 10°	0.81 ± 0.023	32.2 ± 3.6°	0.989 ± 0.002
SS IBI, state A	17 ± 8.8°	0.85 ± 0.017	15 ± 7.4°	0.89 ± 0.014
SS IBI, state B	23 ± 4.9°	0.983 ± 0.005	21.3 ± 3.3°	0.993 ± 0.001
SS IBI, state C	23 ± 4.1°	0.988 ± 0.002	20 ± 4.6°	0.987 ± 0.002
United atom	20 ± 15°	0.820 ± 0.028	32.4 ± 10.1°	0.965 ± 0.046

While in this case, tilt angle and nematic order demonstrated a strong sensitivity to the weighting factors, not all systems or properties will necessarily demonstrate such clear dependence. The ability of the CG potential to reproduce the properties of the target model may certainly be linked to the level of coarse-graining applied (i.e., how many atoms and what type of atoms are lumped together into a single CG bead) and therefore deviations of desired properties may be inherent to the CG model, not the choice of weighting functions. Similarly, as previously discussed in Sec. III C, the fitting of the topological parameters (e.g., bonds, angles, and/or dihedrals) can also influence the measured properties (e.g., R_{chain}), and thus deficiencies caused by these topological parameters are not likely to be fixed via a simple adjustment of weights. However, adjustment of the weights should provide a means to tune potentials that might otherwise be adjusted manually or would need to be derived by mixing potentials from different state points. The use of MS IBI simply provides a clearer, more systematic framework for this process, even if the weights are modified via trial and error.

IV. CONCLUSION

A multi-state extension of the popular IBI method has been proposed. In the proposed MS IBI method, multiple thermodynamic states are used in the derivation of a single, generally applicable potential. For systems with a known potential, it was shown that the MS IBI method was capable of accurately recovering the true, underlying potential, while the SS IBI method was unable to consistently derive a generally applicable potential. Through the coarse-graining of propane, it was shown that MS IBI was able to recover a potential very similar to a previously published single-site model, with excellent reproducibility. Furthermore, potentials derived via MS IBI were shown to better reproduce structural conformations of *n*-dodecane than potentials derived via SS IBI. It was also demonstrated that adjusting the relative weights

given to each target in the optimizations can be used to tune system properties beyond the RDF; in this case, tuning the weights enabled potentials to be derived that provided near perfect agreement between CG and atomistic models when considering the nematic order parameter and tilt angle of an *n*-dodecane monolayer. While pressure, and thermodynamics in general, were not investigated in this work, the standard pressure correction scheme of SS IBI[39] could be trivially applied to MS IBI by calculating the average pressure deviations between all states. As such, the MS IBI stands as an improvement of the typical IBI method, producing more generally applicable potentials that can be tuned to match target properties from experiment or finer-grained simulations.

This improved methodology should be very useful for a host of molecular systems, including, for example, lipid systems, where not only do systems demonstrate structural heterogeneity within a given state point (i.e., different molecular structures in a single system), but properties such as tilt angle, nematic order, area per lipid, etc., need to be tuned in order to match atomistic simulations and experiment.[64–68] Given that the MS IBI approach is also capable of deriving potentials which demonstrate increased levels of transferability than SS IBI, potentials can therefore be derived for complex systems with many unique interactions by examining the individual components separately, reducing the number of simultaneous optimizations that need to be performed. Furthermore, this work presents a method to develop potentials that may allow the examination of phase transitions; in many prior works utilizing SS IBI, different potentials are needed to appropriately model different states, making it difficult to accurately examine the transition between those states.[18,46,69,70] Additionally, given that multi-GPU machines and GPU enabled simulation packages[58,71–74] are becoming more common, the potential derivation process can be performed quite efficiently, even if a large number of targets are needed or a large number of iterations must be undertaken to find appropriate weighting functions.

ACKNOWLEDGMENTS

The authors acknowledge support from Grant No. R01 AR057886-01 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases.

[1]J. N. Israelachvili, D. J. Mitchell, and B. W. Ninham, "Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers," J. Chem. Soc., Faraday Trans. 2 **72**, 1525–1568 (1976).
[2]J. N. Israelachvili, D. J. Mitchell, and B. W. Ninham, "Theory of self-assembly of lipid bilayers and vesicles," Biochem. Biophys. Acta **470**, 185–201 (1977).
[3]I. Muševič and M. Škarabot, "Self-assembly of nematic colloids," Soft Matter **4**, 195 (2008).
[4]K. J. M. Bishop *et al.*, "Nanoscale forces and their uses in self-assembly," Small **5**, 1600–1630 (2009).
[5]C. R. Iacovella, A. S. Keys, and S. C. Glotzer, "Self-assembly of soft-matter quasicrystals and their approximants," Proc. Natl. Acad. Sci. U.S.A. **108**, 20935–20940 (2011).
[6]P. Ziherl and R. D. Kamien, "Maximizing entropy by minimizing area: Towards a new principle of self-organization," J. Phys. Chem. B **105**, 10147–10158 (2001).
[7]E. L. Thomas *et al.*, "Ordered packing arrangements of spherical micelles of diblock copolymers in two and three dimensions," Macromolecules **20**, 2934–2939 (1987).
[8]C. L. Phillips, C. R. Iacovella, and S. C. Glotzer, "Stability of the double gyroid phase to nanoparticle polydispersity in polymer-tethered nanosphere systems," Soft Matter **6**, 1693 (2010).
[9]C. R. Iacovella and S. C. Glotzer, "Complex crystal structures formed by the self-assembly of ditethered nanospheres," Nano Lett. **9**, 1206–1211 (2009).
[10]M. A. Glaser *et al.*, "Soft spheres make more mesophases," Europhys. Lett. **78**, 46004 (2007).
[11]P. Ziherl and R. Kamien, "Soap froths and crystal structures," Phys. Rev. Lett. **85**, 3528–3531 (2000).
[12]R. G. Larson, L. E. Scriven, and H. T. Davis, "Monte Carlo simulation of model amphiphile-oil–water systems," J. Chem. Phys. **83**, 2411 (1985).
[13]S. C. Glotzer *et al.*, "Self-assembly of anisotropic tethered nanoparticle shape amphiphiles," Curr. Opin. Colloid Interface Sci. **10**, 287–295 (2005).
[14]L. Gai *et al.*, "A Wang-Landau study of a lattice model for lipid bilayer self-assembly," J. Chem. Phys. **137**, 144901 (2012).
[15]C. R. Iacovella *et al.*, "Phase diagrams of self-assembled mono-tethered nanospheres from molecular simulation and comparison to surfactants," Langmuir **21**, 9488–9494 (2005).
[16]K. F. Lau and K. A. Dill, "A lattice statistical mechanics model of the conformational and sequence spaces of proteins," Macromolecules **22**, 3986–3997 (1989).
[17]K. A. Maerzke and J. I. Siepmann, "Transferable potentials for phase equilibria-coarse-grain description for linear alkanes," J. Phys. Chem. B **115**, 3452–3465 (2011).
[18]S. J. Marrink *et al.*, "The MARTINI force field: Coarse grained model for biomolecular simulations," J. Phys. Chem. B **111**, 7812–7824 (2007).
[19]M. G. Martin and J. I. Siepmann, "Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes," J. Phys. Chem. B **102**, 2569–2577 (1998).
[20]W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," J. Am. Chem. Soc. **118**, 11225–11236 (1996).
[21]C. Oostenbrink *et al.*, "A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6," J. Comput. Chem. **25**, 1656–1676 (2004).
[22]K. Vanommeslaeghe *et al.*, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," J. Comput. Chem. **31**, 671–690 (2009).
[23]J. Wang *et al.*, "Development and testing of a general amber force field," J. Comput. Chem. **25**, 1157–1174 (2004).
[24]C. McCabe *et al.*, "Multiscale simulation of the synthesis, assembly and properties of nanostructured organic/inorganic hybrid materials," J. Comput. Theory Nanosci. **1**, 265–279 (2004).
[25]H. Liu *et al.*, "Multiscale simulation study on the curing reaction and the network structure in a typical epoxy system," Macromolecules **44**, 8650–8660 (2011).
[26]C. Peter and K. Kremer, "Multiscale simulation of soft matter systems – From the atomistic to the coarse-grained level and back," Soft Matter **5**, 4357 (2009).
[27]N. di Pasquale, D. Marchisio, and P. Carbone, "Mixing atoms and coarse-grained beads in modelling polymer melts," J. Chem. Phys. **137**, 164111 (2012).
[28]B. Ensing *et al.*, "Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics," J. Chem. Theory Comput. **3**, 1100–1105 (2007).
[29]E. Lidorikis *et al.*, "Coupling length scales for multiscale atomistics-continuum simulations: Atomistically induced stress distributions in Si/Si3N4 nanopixels," Phys. Rev. Lett. **87**, 086104 (2001).
[30]J. Michel, M. Orsi, and J. W. Essex, "Prediction of partition coefficients by multiscale hybrid atomic-level/coarse-grain simulations," J. Phys. Chem. B **112**, 657–660 (2008).
[31]M. Praprotnik, L. D. Site, and K. Kremer, "Multiscale simulation of soft matter: From scale bridging to adaptive resolution," Ann. Rev. Phys. Chem. **59**, 545–571 (2008).
[32]A. J. Rzepiela *et al.*, "Hybrid simulations: Combining atomistic and coarse-grained force fields using virtual sites," Phys. Chem. Chem. Phys. **13**, 10437–10448 (2011).
[33]T. Werder, J. H. Walther, and P. Koumoutsakos, "Hybrid atomistic–continuum method for the simulation of dense fluid flows," J. Comput. Phys. **205**, 373–390 (2005).

[34]A. P. Heath, L. E. Kavraki, and C. Clementi, "From coarse-grain to all-atom: Toward multiscale analysis of protein landscapes," Proteins **68**, 646–661 (2007).
[35]B. Hess, C. Holm, and N. van der Vegt, "Osmotic coefficients of atomistic NaCl (aq) force fields," J. Chem. Phys. **124**, 164509 (2006).
[36]P. Liu *et al.*, "Reconstructing atomistic detail for coarse-grained models with resolution exchange," J. Chem. Phys. **129**, 114103 (2008).
[37]A. J. Rzepiela *et al.*, "Software news and update reconstruction of atomistic details from coarse-grained structures," J. Comput. Chem. **31**, 1333–1343 (2010).
[38]E. Lyman, F. M. Ytreberg, and D. M. Zuckerman, "Resolution exchange simulation," Phys. Rev. Lett. **96**, 028105 (2006).
[39]D. Reith, M. Putz, and F. Muller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations," J. Comput. Chem. **24**, 1624–1636 (2003).
[40]F. Ercolessi and J. B. Adams, "Interatomic potentials from first-principles calculations: The force matching method," Europhys. Lett. **26**, 583–588 (1994).
[41]S. Izvekov and G. A. Voth, "A multiscale coarse-graining method for biomolecular systems," J. Phys. Chem. B **109**, 2469–2473 (2005).
[42]M. S. Shell, "The relative entropy is fundamental to multiscale and inverse thermodynamic problems," J. Chem. Phys. **129**, 144108 (2008).
[43]A. Chaimovich and M. S. Shell, "Coarse-graining errors and numerical optimization using a relative entropy framework," J. Chem. Phys. **134**, 094112 (2011).
[44]C. R. Iacovella *et al.*, "Pair interaction potentials of colloids by extrapolation of confocal microscopy measurements of collective suspension structure," J. Chem. Phys. **133**, 164903 (2010).
[45]B. Bozorgui *et al.*, "Fluctuation-driven anisotropic assembly in nanoscale systems," Nano Lett. **13**, 2732–2737 (2013).
[46]K. R. Hadley and C. McCabe, "A coarse-grained model for amorphous and crystalline fatty acids," J. Chem. Phys. **132**, 134505 (2010).
[47]R. Faller, "Automatic coarse graining of polymers," Polymer **45**, 3869–3876 (2004).
[48]C.-C. Fu *et al.*, "A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties," J. Chem. Phys. **137**, 164106 (2012).
[49]H.-J. Qian *et al.*, "Temperature-transferable coarse-grained potentials for ethylbenzene, polystyrene, and their mixtures," Macromolecules **41**, 9919–9929 (2008).
[50]B. Bayramoglu and R. Faller, "Coarse-grained modeling of polystyrene in various environments by iterative Boltzmann inversion," Macromolecules **45**, 9205–9219 (2012).
[51]B. Bayramoglu and R. Faller, "Modeling of polystyrene under confinement: Exploring the limits of iterative Boltzmann inversion," Macromolecules **46**, 7957–7976 (2013).
[52]P. Ganguly and N. F. A. van der Vegt, "Representability and transferability of Kirkwood − Buff iterative Boltzmann inversion models for multicomponent aqueous systems," J. Chem. Theory Comput. **9**, 5247–5226 (2013).
[53]E. R. Chan *et al.*, "Coarse-grained force field for simulating polymer-tethered silsesquioxane self-assembly in solution," J. Chem. Phys. **127**, 114102 (2007).
[54]G. Milano, S. Goudeau, and F. Müller-Plathe, "Multicentered Gaussian-based potentials for coarse-grained polymer simulations: Linking atomistic and mesoscopic scales," J. Polym. Sci., Part B: Polym. Phys. **43**, 871–885 (2005).
[55]K. R. Hadley and C. McCabe, "A structurally relevant coarse-grained model for cholesterol," Biophys. J. **99**, 2896–2905 (2010).
[56]J. L. Rivera, G. K. Jennings, and C. McCabe, "Examining the frictional forces between mixed hydrophobic-hydrophilic alkylsilane monolayers," J. Chem. Phys. **136**, 244701 (2012).
[57]See supplementary material at http://dx.doi.org/10.1063/1.4880555 for derivation of n-Dodecane pair and bonded potentials and associated RDFs.
[58]J. A. Anderson, C. D. Lorenz, and A. Travesset, "General purpose molecular dynamics simulations fully implemented on graphics processing units," J. Comput. Phys. **227**, 5342–5359 (2008).
[59]See http://codeblue.umich.edu/hoomd-blue for the description, documentation, and installation instructions for HOOMD-Blue.
[60]J. D. Weeks, D. Chandler, and H. C. Andersen, "Role of repulsive forces in determining the equilibrium structure of simple liquids," J. Chem. Phys. **54**, 5237–5247 (1971).
[61]Q. Pu *et al.*, "Molecular simulations of stretching gold nanowires in solvents," Nanotechnology **18**, 424007 (2007).
[62]K. R. Hadley and C. McCabe, "A simulation study of the self-assembly of coarse-grained skin lipids," Soft Matter **8**, 4802–4814 (2012).
[63]A. S. Keys, C. R. Iacovella, and S. C. Glotzer, "Characterizing complex particle morphologies through shape matching: Descriptors, applications, and algorithms," J. Comput. Phys. **230**, 6438–6463 (2011).
[64]H. Heller, M. Schaefer, and K. Schulten, "Molecular-dynamics simualtion of a bilayer of 200 lipids in the gel and in the liquid-crystal phases," J. Phys. Chem. **97**, 8343–8360 (1993).
[65]E. Egberts, S. J. Marrink, and H. J. C. Berendsen, "Molecular-dynamics simulation of a phospholipid membrane," Eur. Biophys. J. **22**, 423–436 (1994).
[66]L. Saiz and M. L. Klein, "Computer simulation studies of model biological membranes," Acc. Chem. Res. **35**, 482–489 (2002).
[67]H. L. Scott, "Modeling the lipid component of membranes," Curr. Opin. Struct. Biol. **12**, 495–502 (2002).
[68]M. Venturoli *et al.*, "Mesoscopic models of biological membranes," Phys. Rep.-Rev. Sec. Phys. Lett. **437**, 1–54 (2006).
[69]T. Vettorel and H. Meyer, "Coarse graining of short polythylene chains for studying polymer crystallization," J. Chem. Theory Comput. **2**, 616–629 (2006).
[70]G. D'Adamo, A. Pelissetto, and C. Pierleoni, "Coarse-graining strategies in polymer solutions," Soft Matter **8**, 5151–5167 (2012).
[71]B. Hess *et al.*, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," J. Chem. Theory Comput. **4**, 435–447 (2008).
[72]S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," J. Comput. Phys. **117**, 1–19 (1995).
[73]W. M. Brown *et al.*, "Implementing molecular dynamics on hybrid high performance computers – Short range forces," Comput. Phys. Commun. **182**, 898–911 (2011).
[74]W. M. Brown *et al.*, "Implementing molecular dynamics on hybrid high performance computers – Particle–particle particle-mesh," Comput. Phys. Commun. **183**, 449–459 (2012).

Supplemental Material: Derivation of Coarse-grained Potentials via Multistate Iterative Boltzmann Inversion

*Timothy C. Moore[1,3], Christopher R. Iacovella[1,3], Clare McCabe[1,2,3]**

[1]*Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN 37235*

[2]*Department of Chemistry, Vanderbilt University, Nashville, TN 37235*

[3]*Vanderbilt Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, TN 37235*

**Corresponding Author. Email: c.mccabe@vanderbilt.edu*

_n_-Dodecane optimizations

Potentials were derived for *n*-dodecane using both MS IBI and SS IBI. As described in the main text, the following states were used: state A, 298 K, 1.04 g/mL; state B, 298 K, 0.74 g/mL; and state C, 370 K, 0.55 g/mL; the damping values used were $\alpha_A(0) = 0.5$, $\alpha_B(0) = 0.7$, and $\alpha_C(0) = 0.5$. State B, was was used in the SS IBI fitting, with a damping factor of $\alpha_B(0) = 0.7$. In both cases, good agreement between trial RDFs and target RDFs was found at each state, as shown in Figures S1-3 (MS IBI) and Figures S4-6 (SS-IBI). The two-point smoothing function described in the methods section was applied to each potential at the end of 200 iterations.

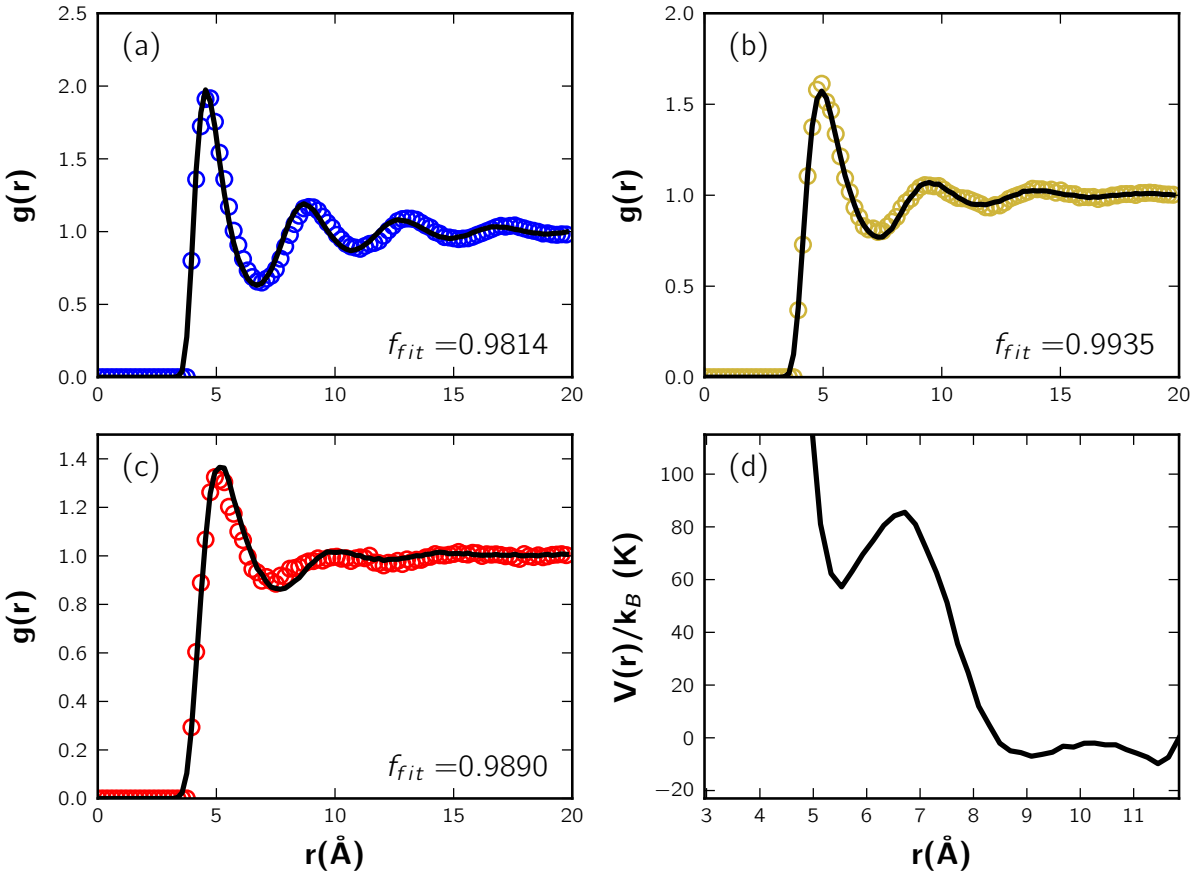

Figure S1. End-end bead RDFs and potential derived via MS IBI, after 200 iterations.

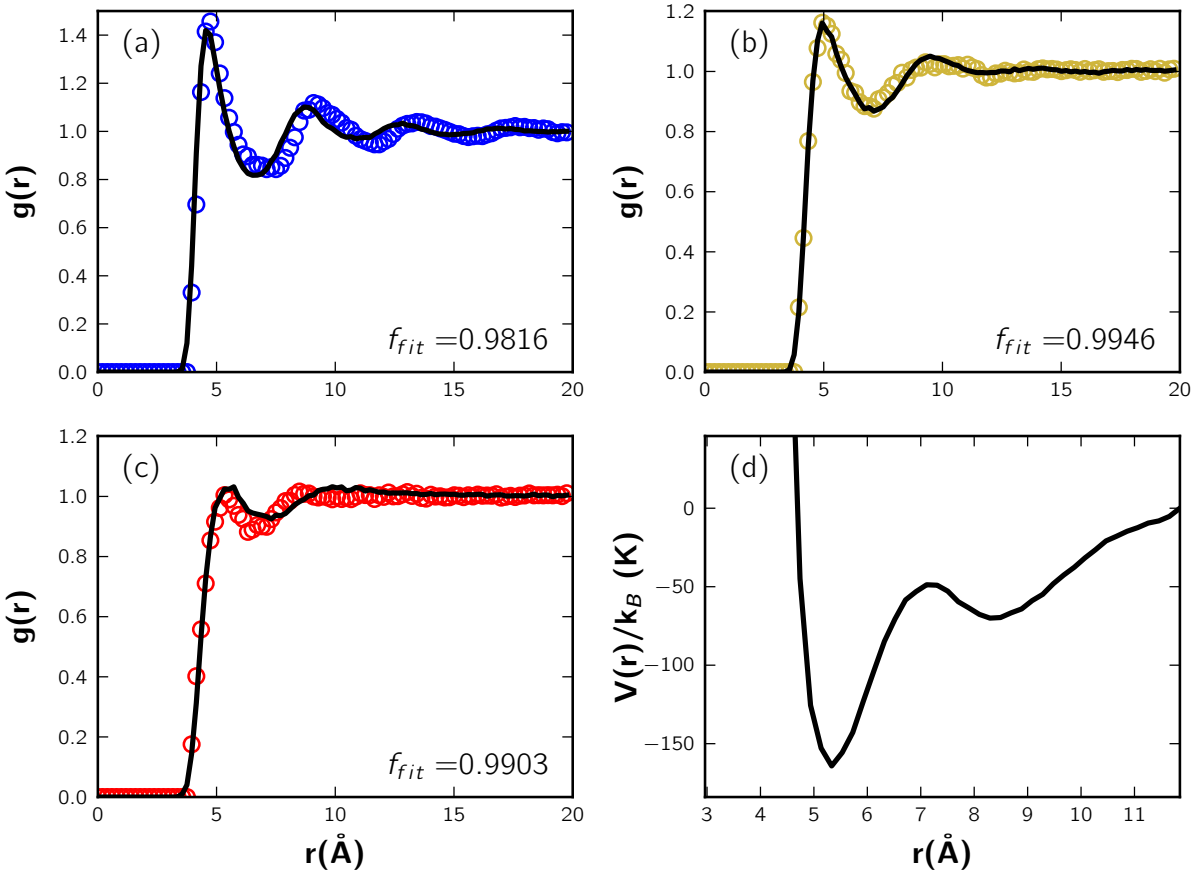

Figure S2. Middle-middle bead RDFs and potential derived via MS IBI, after 200 iterations.

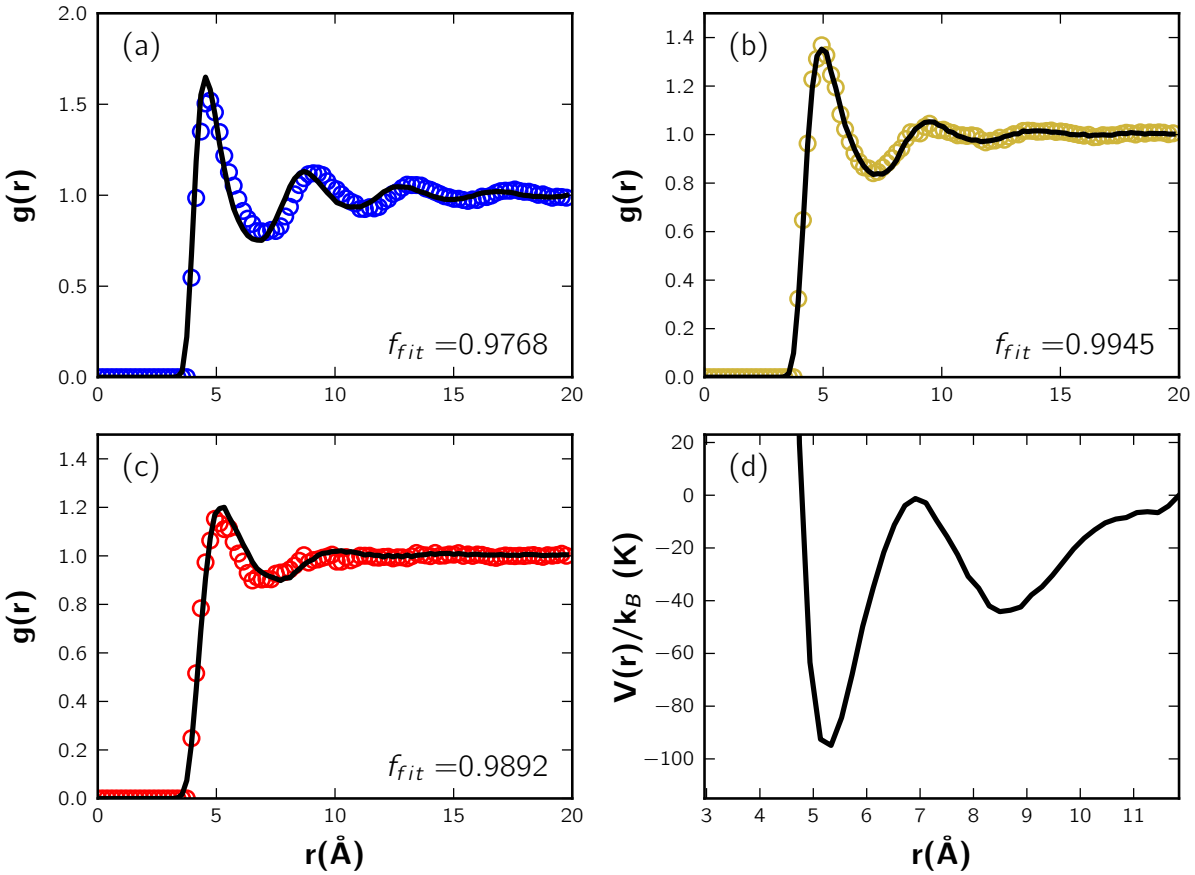

Figure S2. Middle-end bead RDFs and potential derived via MS IBI, after 200 iterations.

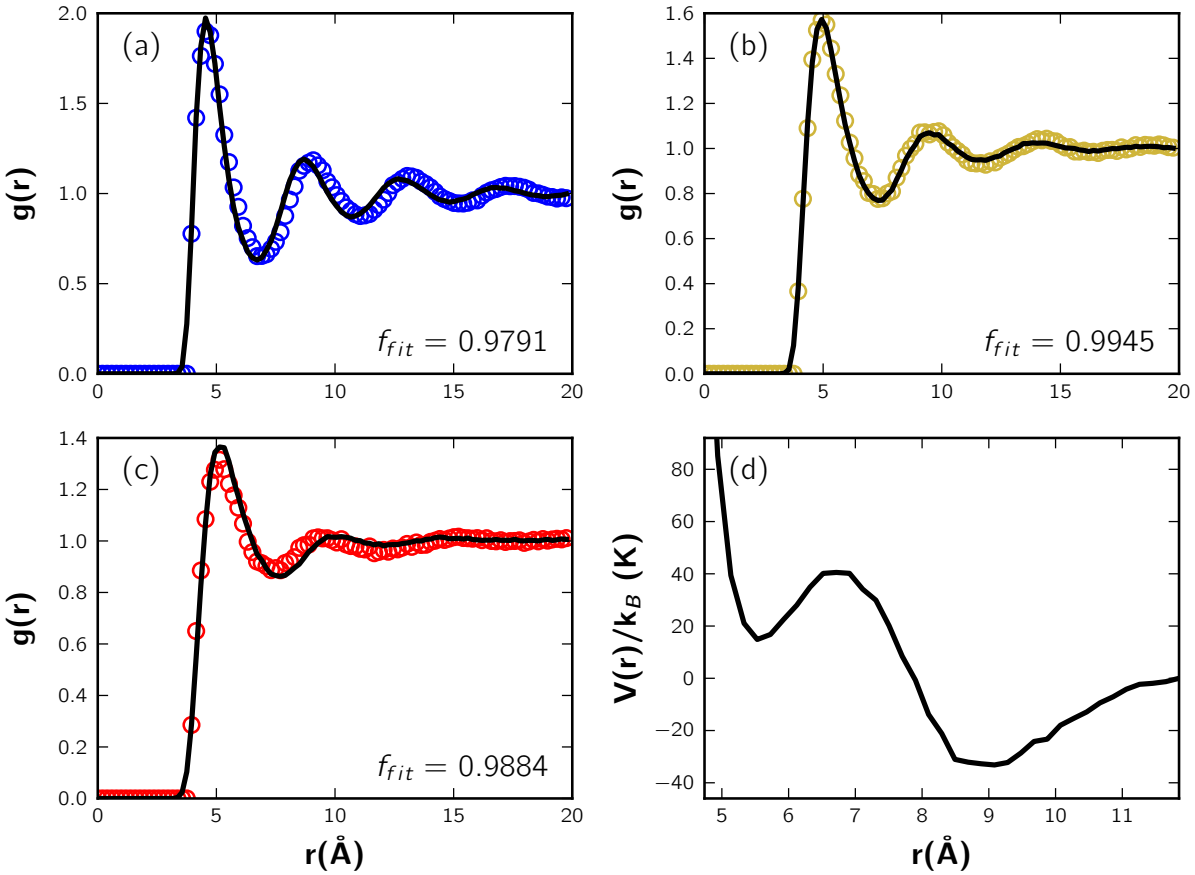

Figure S3. End-end bead RDFs and potential derived via SS IBI, after 200 iterations.

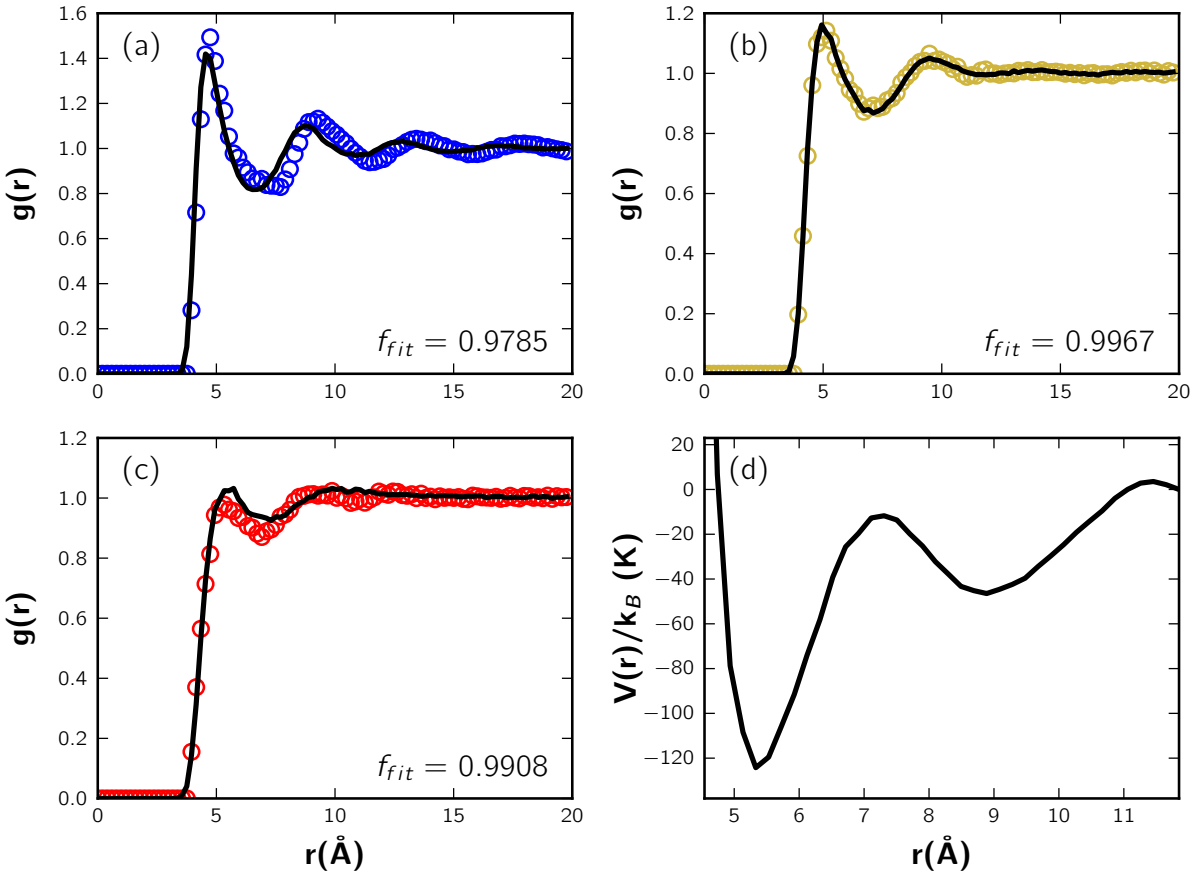

Figure S4. Middle-middle bead RDFs and potential derived via SS IBI, after 200 iterations.

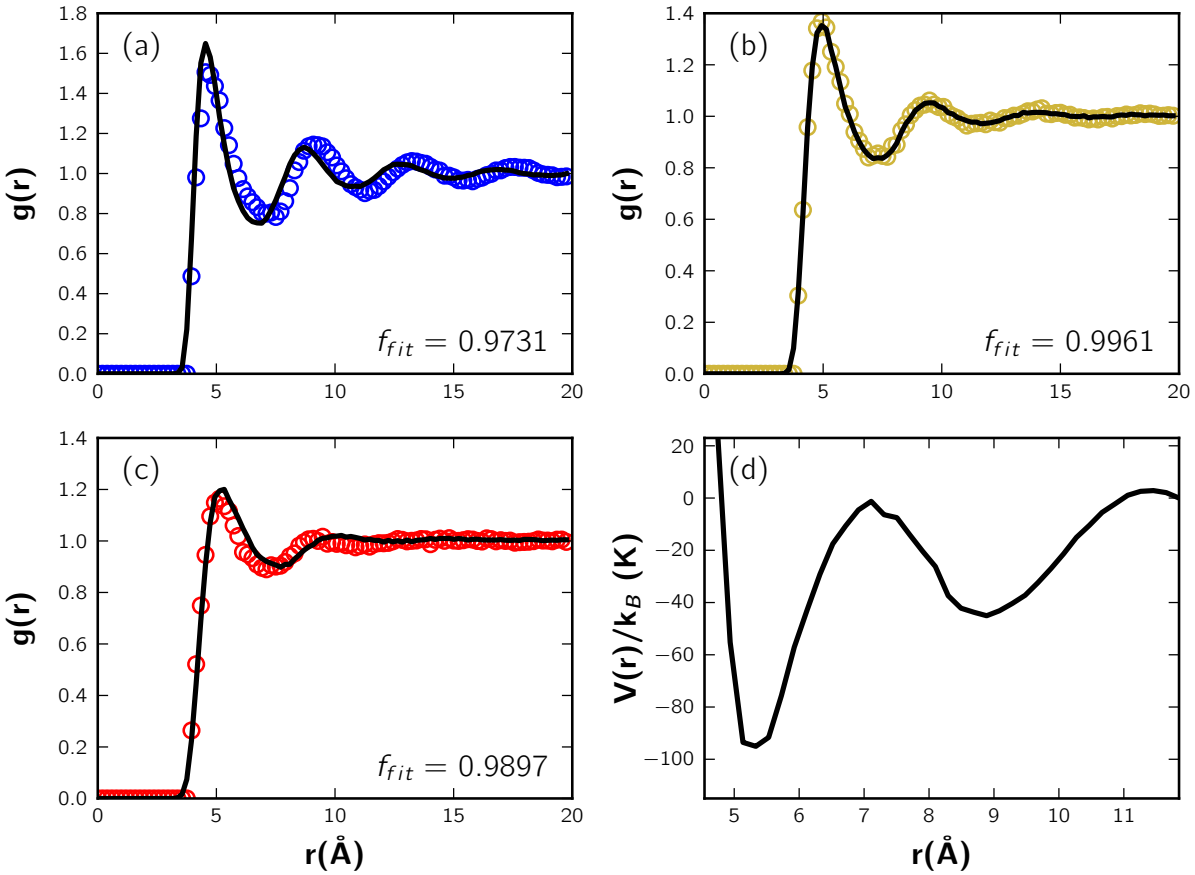

Figure S6. Middle-end bead RDFs and potential derived via SS IBI, after 200 iterations.

Bond and angle fittings

As mentioned in the methods section of the main text, a single set of bond and angle parameters was used in all CG *n*-dodecane simulations (i.e., bulk fluid and monolayer simulations). These parameters were calculated by fitting a Gaussian curve to the bond and angle distributions sampled in the atomistic trajectory at state B (described above and in the main text), after the atomistic trajectory was mapped to the CG level. Gaussians were fit to the data points using a least squares fitting algorithm. Shown below are the bond and angle distributions for all states. Note that there are two unique bond types (i.e., END-MID and MID-MID), but they are shown on the same plot for each state. Additionally, note that only a single Gaussian was used for each data set.

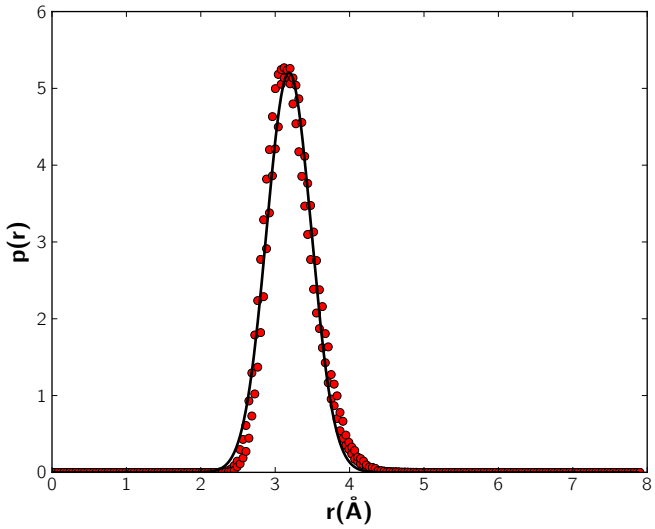

Figure S7. Bond length distribution for *n*-dodecane simulated at state A using the TraPPE-UA model. The trajectory was mapped to the CG level before calculating the distribution. Red dots represent the target data, the solid black line represents the Gaussian fit.

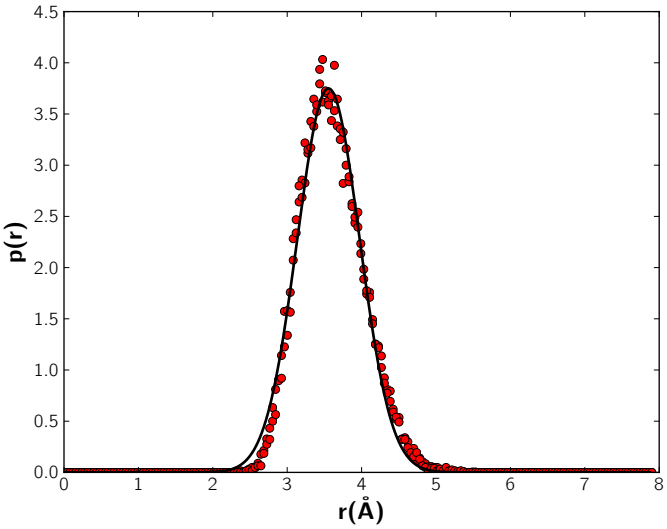

Figure S8. Bond length distribution for *n*-dodecane simulated at state B using the TraPPE-UA model. The trajectory was mapped to the CG level before calculating the distribution. Red dots represent the target data, the solid black line represents the Gaussian fit.

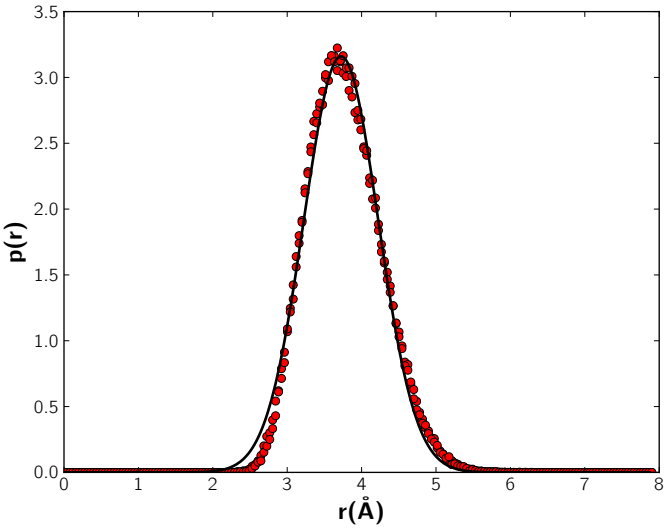

Figure S9. Bond length distribution for *n*-dodecane simulated at state C using the TraPPE-UA model. The trajectory was mapped to the CG level before calculating the distribution. Red dots represent the target data, the solid black line represents the Gaussian fit.

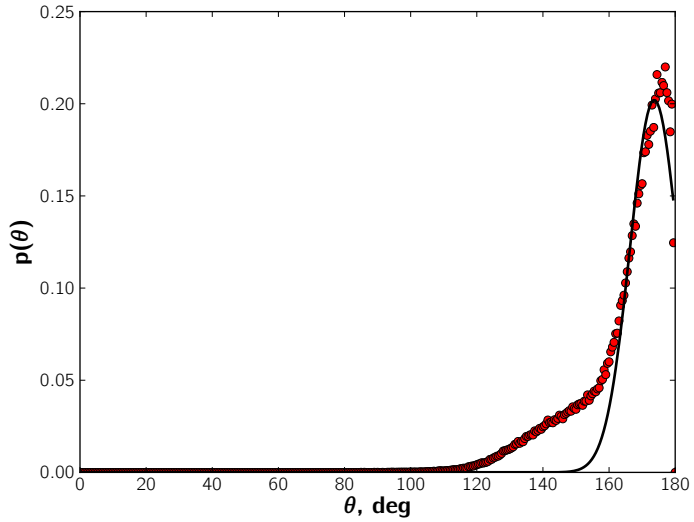

Figure S10. Angle distribution for *n*-dodecane simulated at state A using the TraPPE-UA model. The trajectory was mapped to the CG level before calculating the distribution. Red dots represent the target data, the solid black line represents the Gaussian fit.

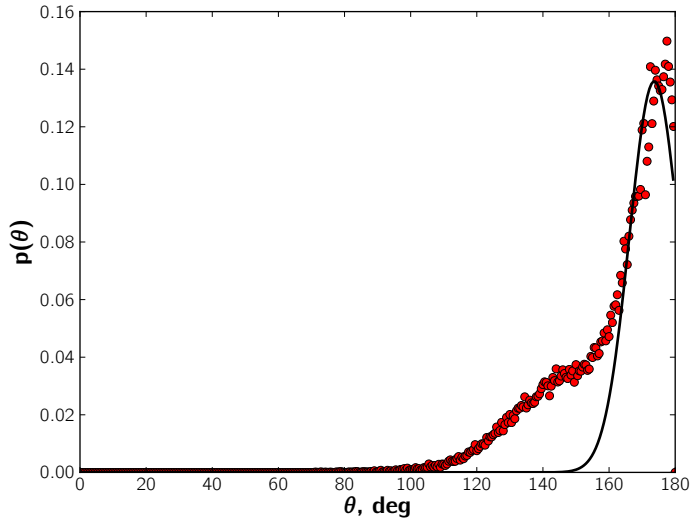

Figure S11. Angle distribution for *n*-dodecane simulated at state B using the TraPPE-UA model. The trajectory was mapped to the CG level before calculating the distribution. Red dots represent the target data, the solid black line represents the Gaussian fit.

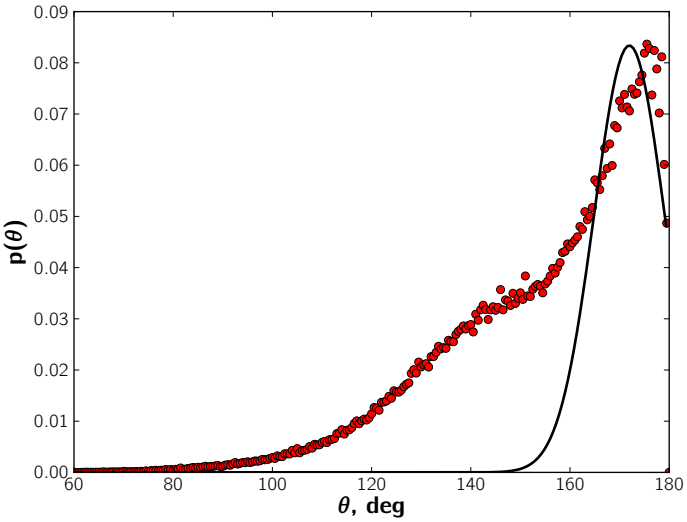

Figure S12. Angle distribution for *n*-dodecane simulated at state C using the TraPPE-UA model. The trajectory was mapped to the CG level before calculating the distribution. Red dots represent the target data, the solid black line represents the Gaussian fit.