van der Waals perspective on coarse-graining: Progress towards solving representability and transferability problems

Nicholas J.H. Dunn,[†,¶] Thomas T. Foley,[‡,¶] and William G. Noid[*,†]

†*Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802*

‡*Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802*

¶*Equally contributing author*

E-mail: wnoid@chem.psu.edu

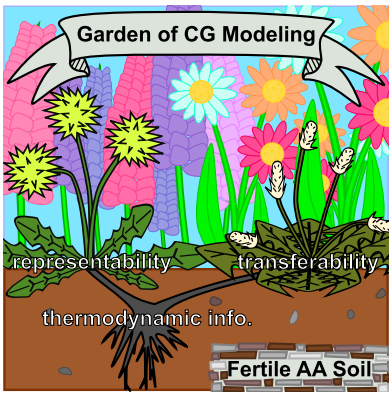
Conspectus

Low-resolution coarse-grained (CG) models provide the necessary efficiency for simulating phenomena that are inaccessible to more detailed models. However, in order to realize their considerable promise, CG models must accurately describe the relevant physical forces and provide useful predictions. By formally integrating out the unnecessary details from an all-atom (AA) model, "bottom-up" approaches can, at least in principle, quantitatively reproduce the structural and thermodynamic properties of the AA model that are observable at the CG resolution. In practice, though, bottom-up approaches only approximate this "exact coarse-graining" procedure. The resulting models typically reproduce the intermolecular structure of AA models at a single thermodynamic state point, but often describe other state points less accurately and, moreover, tend to provide a poor description of thermodynamic properties. These

two limitations have been coined the "transferability" and "representability" problems, respectively. Perhaps, the simplest and most commonly discussed manifestation of the representability problem regards the tendency of structure-based CG models to dramatically over-estimate the pressure. Furthermore, when these models are adjusted to reproduce the pressure, they provide a poor description of the compressibility. More generally, it is sometimes suggested that CG models are fundamentally incapable of reproducing both structural and thermodynamic properties. After all, there is no such thing as a "free lunch" - any significant gain in computational efficiency should come at the cost of significant model limitations.

At least in the case of structural and thermodynamic properties, though, we optimistically propose that this may be a false dichotomy. Accordingly, we have recently re-examined the "exact coarse-graining" procedure and investigated the intrinsic consequences of representing an AA model in reduced resolution. These studies clarify the origin and inter-relationship of representability and transferability problems. Both arise as consequences of transferring thermodynamic information from the high resolution configuration space and encoding this information into the many-body potential of mean force (PMF), i.e., the potential that emerges from an exact coarse-graining procedure. At least in principle, both representability and transferability problems can be resolved by properly addressing this thermodynamic information. In particular, we have demonstrated that "pressure-matching" provides a practical and rigorous means for addressing the density-dependence of the PMF. The resulting bottom-up models accurately reproduce the structure, equilibrium density, compressibility, and pressure equation of state for AA models of molecular liquids. Additionally, we have extended this approach to develop transferable potentials that provide similar accuracy for heptane-toluene mixtures. Moreover, these potentials provide predictive accuracy for modeling concentrations that were not considered in their parameterization. More generally, this work suggests a "van der Waals" perspective on coarse-graining, in which conventional structure-based methods accurately describe the configuration-dependence of the PMF, while independent variational principles infer the thermo-

dynamic information that is necessary to resolve representability and transferability problems.



1 Introduction

Low resolution coarse-grained (CG) models play an important and rapidly growing role in science.[1,2] By eliminating unnecessary details, CG models provide the necessary efficiency for simulating length- and time-scales that remain inaccessible to more detailed models.[3] CG models also empower more systematic investigations of the relevant experimental conditions, while simultaneously providing superior statistical precision in simulated quantities. Furthermore, CG models more effectively harness the intellectual horsepower of researchers by focusing attention on the essential details of a particular phenomenon, which atomically detailed models can easily obscure.[4,5]

Historically, CG models have been extensively employed for investigating the emergent consequences of basic physical principles.[6] More recently, though, many coarse-graining approaches have been developed for modeling specific systems.[7] By formally integrating out unnecessary atomic details, "bottom-up" approaches can, at least in principle, quantitatively reproduce the structural and thermodynamic properties of a high resolution model that can be observed at the resolution of the CG model, although thermodynamic properties require careful consideration.[3,6] Of course, this "exact coarse-graining" procedure cannot be accomplished in practice. Rather, bottom-up models are often parameterized to accurately describe the structure of a high resolution model at a single thermodynamic state point.[8] Unfortunately, the resulting models often prove accurate over a relatively limited range of thermodynamic conditions and, moreover, tend to provide a surprisingly poor description of thermodynamic properties.[7] These difficulties are termed "transferability" and "representability" problems, respectively.

Transferability problems are not surprising. CG potentials are constructed to incorporate the effects of atoms that have been eliminated from the CG model. These effects will certainly vary with thermodynamic state point. Thus, one intuitively expects that CG potentials should depend upon thermodynamic conditions. Indeed, many previous studies have documented the sensitivity of bottom-up potentials to variations in state point.[9–19]

Consequently, one expects that any approximate potential will accurately describe these effects only over a relatively limited range of thermodynamic conditions.

Representability problems are more subtle. While recent studies introduced the term to describe thermodynamic inconsistencies in CG models,[20,21] representability problems are related to inconsistencies observed much earlier for effective potentials employed in liquid state theories.[22,23] Perhaps the simplest and most commonly discussed representability problem regards the pressure-volume behavior of structure-based CG models. Indeed, many studies have observed that structure-based CG models generate unrealistically high pressures.[24] For instance, under ambient conditions, structure-based CG models overestimate the internal pressure of liquid water by almost four orders of magnitude.[21,25] Moreover, when these models are modified to reproduce the pressure, they then provide a poor description of the isothermal compressibility.[26] Similarly, previous studies have also demonstrated that CG models poorly describe the energetic and entropic contributions to free energy differences.[27–29] It has been suggested that representability problems reflect fundamental limitations of state-point-dependent effective potentials and that, more simply, CG models cannot accurately describe multiple conflicting observables, such as the pressure and the compressibility.[21,30] Alternatively, it has been proposed that representability problems may be resolved by considering the effects of the missing atomic degrees of freedom upon the CG representation of thermodynamic observables.[7,31] In particular, Guenza and coworkers have demonstrated the importance of these effects for low resolution polymer models.[32–34]

This account summarizes our recent studies of representability and transferability challenges.[35–37] We have adopted the optimistic hypothesis that both challenges can be resolved by carefully considering exact coarse-graining and the intrinsic consequences of representing a system in reduced detail. Our analysis clarifies both the origin and inter-relation of representability and transferability problems. Moreover, our work demonstrates an extended ensemble pressure-matching approach[38,39] for determining transferable potentials that accurately model the structure, pressure, and compressibility of molecular liquids in practice.

2 Exact coarse-graining

2.1 Atomic model

We first consider the canonical ensemble for an all-atom (AA) model that represents a system with n atoms labelled $i = 1, \ldots, n$ in a volume, V, at a temperature, T.[40] We indicate the atomic configuration, $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_n)$. The atoms interact according to a conservative potential, $u(\mathbf{r}; V)$, that generates a force, $\mathbf{f}_i(\mathbf{r}; V)$, on each atom i, as well as a force on the volume, i.e., the fluctuating internal pressure:

$$p_{\text{int}}(\mathbf{r}; V, T) = nk_BT/V - (\partial u(\mathbf{r}; V)/\partial V)_{\hat{\mathbf{r}}} \,. \tag{1}$$

The canonical ensemble average of $p_{\text{int}}(\mathbf{r}; V, T)$ equals the thermodynamic internal pressure of the AA model, $p_{\text{int}}(V, T)$.

The first term in Eq. (**??**) describes the kinetic, i.e., ideal, contribution to the pressure. The second term defines the instantaneous excess (xs) pressure, $p_{\text{xs}}(\mathbf{r}; V) = -(\partial u(\mathbf{r}; V)/\partial V)_{\hat{\mathbf{r}}}$, while $\hat{\mathbf{r}} = \left(V^{-1/3}\mathbf{r}_1, \ldots, V^{-1/3}\mathbf{r}_n\right)$ denotes the "scaled configuration." This contribution is often calculated from the virial expression:

$$p_{\text{xs}}(\mathbf{r}; V) = \frac{1}{3V} \sum_i \mathbf{f}_i(\mathbf{r}; V) \cdot \mathbf{r}_i = \frac{1}{3V} \sum_{(i,j)} f_{2;ij}(r_{ij}) r_{ij}, \tag{2}$$

where the second sum is performed over all intra- and inter-molecular pairs (i, j) that are separated by a distance r_{ij} and interact with a force of magnitude $f_{2;ij}(r_{ij})$. Both expressions for $p_{\text{xs}}(\mathbf{r}; V)$ assume that the atomic potential does not explicitly depend upon the volume, i.e., $(\partial u/\partial V)_{\mathbf{r}} = 0$. The second expression also assumes that the nonbonded potential is pair-additive. While angle-dependent bonded potentials do not contribute to the virial, more complex non-bonded interactions may introduce additional contributions.[36,40]

Finally, we consider the total differential describing variations in the atomic potential,

i.e., work:

$$\mathrm{d}u(\mathbf{r};V) = -\sum_i \mathbf{f}_i(\mathbf{r};V)\cdot(\mathrm{d}\mathbf{r}_i)_V - p_{\mathrm{xs}}(\mathbf{r};V)\mathrm{d}V, \tag{3}$$

where $(\mathrm{d}\mathbf{r}_i)_V = V^{1/3}\mathrm{d}\hat{\mathbf{r}}_i$. The first term in Eq. (**??**) quantifies changes in potential energy due to configuration changes at constant volume. The second term quantifies changes in potential energy due to isotropic compression or expansion.

2.2 Coarse-grained model

We next consider the canonical ensemble (at the same V and T) for a CG model that describes the same system with N "sites" that are labelled $I = 1,\ldots,N$. We indicate the CG configuration, $\mathbf{R} = (\mathbf{R}_1,\ldots,\mathbf{R}_N)$. The sites interact according to a conservative potential, $U(\mathbf{R},V)$, that generates a force, $\mathbf{F}_I(\mathbf{R},V)$, on each site I, as well as a fluctuating internal pressure:

$$P_{\mathrm{int}}(\mathbf{R},V;T;U) = Nk_BT/V - (\partial U(\mathbf{R},V)/\partial V)_{\hat{\mathbf{R}}}\,. \tag{4}$$

The canonical ensemble average of $P_{\mathrm{int}}(\mathbf{R},V;T;U)$ equals the thermodynamic internal pressure of the CG model, $P_{\mathrm{int}}(V,T;U)$.

As above, the first term in Eq. (**??**) describes the ideal contribution for the $N \leq n$ remaining CG particles. The second term defines the instantaneous excess pressure of the CG model, $P_{\mathrm{xs}}(\mathbf{R},V) = -(\partial U(\mathbf{R},V)/\partial V)_{\hat{\mathbf{R}}}$, while $\hat{\mathbf{R}} = (V^{-1/3}\mathbf{R}_1,\ldots,V^{-1/3}\mathbf{R}_N)$ denotes the scaled CG configuration. Our objective is to parameterize U for accurately describing the structural and thermodynamic properties of the atomic model.

2.3 The many-body Potential of Mean Force

In order to relate the AA and CG models, we introduce a mapping, $\mathbf{M}$, that determines the CG configuration as a function of the AA configuration: $\mathbf{R} = \mathbf{M}(\mathbf{r})$.[41] For simplicity, we

assume the mapping associates the CG sites with the mass centers of disjoint atomic groups. The central quantity in our analysis is the many-body potential of mean force (PMF), which is the effective potential that results from "exact coarse-graining" in the canonical ensemble:

$$\exp\left[-W(\mathbf{R};V,T)/k_BT\right] = V^{-(n-N)} \int_V \mathrm{d}\mathbf{r}\, \exp[-u(\mathbf{r};V)/k_BT]\, \delta\left(\mathbf{M}(\mathbf{r}) - \mathbf{R}\right), \tag{5}$$

where the integral is performed over the volume-dependent configuration space.[42–44] The PMF is the appropriate potential for ensuring that the CG model samples configurations according to the probability implied by the atomistic model and the mapping at the given V and T.[41] Moreover, the PMF encodes all information and thermodynamic properties that are observable at the resolution of the CG model. In particular, the PMF ensures that the CG model reproduces the excess free energy of the AA model:

$$\frac{1}{V^N}\int_V \mathrm{d}\mathbf{R}\, \exp[-W(\mathbf{R};V,T)/k_BT] = \frac{1}{V^n}\int_V \mathrm{d}\mathbf{r}\, \exp[-u(\mathbf{r};V)/k_BT]. \tag{6}$$

2.4 Energetic and entropic contributions

The PMF is not a conventional potential, but rather a free energy that depends upon both the configuration and also the thermodynamic state.[6] In collaboration with the Shell group,[35] we have examined the thermodynamic character of the PMF and, in particular, derived its energetic, U_W, and entropic, S_W, components:

$$W(\mathbf{R};V,T) = U_W(\mathbf{R};V,T) - TS_W(\mathbf{R};V,T) \tag{7}$$

$$U_W(\mathbf{R};V,T) \equiv \langle u(\mathbf{r};V)\rangle_{\mathbf{R};V,T} \tag{8}$$

$$S_W(\mathbf{R};V,T) \equiv \left\langle -k_B \ln\left[\Omega_1 \overline{p}_{r|R}(\mathbf{r}|\mathbf{R};V,T)\right]\right\rangle_{\mathbf{R};V,T}, \tag{9}$$

where $\Omega_1 = V^{n-N}$ is the volume element of atomic configurations $\mathbf{r}$ that map to $\mathbf{R}$, $\overline{p}_{r|R}(\mathbf{r}|\mathbf{R};V,T)$ is the conditioned probability density for AA configurations $\mathbf{r}$ satisfying $\mathbf{M}(\mathbf{r}) = \mathbf{R}$, and the

subscripted angular brackets indicate corresponding conditioned canonical averages.

The energetic contribution to the PMF, $U_W(\mathbf{R};V,T)$, is simply the conditioned average of the atomic potential for the atomic configurations that map to $\mathbf{R}$. This energetic contribution generates forces that bias the CG model to sample low-energy configurations. The entropic contribution, $S_W(\mathbf{R};V,T)$, quantifies the excess entropy that is stored in the Boltzmann distribution of atomic configurations that map to $\mathbf{R}$. Thus, S_W quantifies the information about the atomic distribution that is "lost" when viewing this distribution at the CG resolution. By the Gibbs inequality,[45] $-TS_W \geq 0$, and only vanishes when $\overline{p}_{r|R} = \Omega_1^{-1}$, i.e., when all atomic configurations that map to $\mathbf{R}$ have equal Boltzmann weight. Consequently, $-TS_W$ generates forces that bias the CG model to sample high-entropy configurations, i.e., CG configurations for which the underlying atomic Boltzmann distribution, $\overline{p}_{r|R}$, is more uniform.

This simple decomposition is fundamentally important for representing thermodynamic properties with CG models. Since W is a free energy and incorporates an entropic component, W cannot be directly employed to estimate atomic energies. Similarly, the configurational entropy of the atomic model cannot be directly estimated from the configuration distribution sampled by the CG model.[46] Nevertheless, at least in principle, U_W and S_W can be employed to quantify the excess energy and excess entropy, respectively, of the atomic model, albeit at the resolution of the CG model.

In order to illustrate these considerations, we analytically derived the exact PMF for the Gaussian Network model (GNM) as a function of the CG resolution.[35] The GNM describes protein fluctuations away from an equilibrium structure with a system of linear springs between nearby residues.[47] For each of 7 proteins, we constructed a high resolution GNM that explicitly represented 120 α carbons of the protein. For each protein, we determined W for a series of N-site CG models in which we mapped each consecutive $120/N$ α carbons to their mass center.

Figure 1 illustrates the impact of resolution upon W, U_W, S_W, and s_R, i.e., the excess

entropy present in the configuration space. In the absence of coarse-graining, i.e., $N = 120$, the PMF is simply the atomic potential, $W = u$, $S_W = 0$, and the excess entropy is stored in the atomic configurational distribution. Because U_W is simply a conditioned average of the atomic potential, its average magnitude does not vary with coarsening, as indicated by the dashed horizontal line. However, with successive coarsening, configurational entropy and, equivalently, information is eliminated from the atomistic configuration space. The excess entropy is transferred into S_W, which results in a systematic increase in W with coarsening. In the extreme limit of coarse-graining, $N \to 0$, $s_R \to 0$, and the PMF becomes the configuration-independent, excess Helmholtz potential of the atomic model. In this limit, U_W and S_W become the thermodynamic excess energy and excess entropy, respectively.

2.5 Variation in the PMF

Further insight into representability and transferability issues can be gleaned from the total differential of the PMF:

$$\mathrm{d}W(\mathbf{R};V,T) = -\sum_I \overline{\mathbf{f}}_I(\mathbf{R};V,T)\cdot(\mathrm{d}\mathbf{R}_I)_V - \overline{p}_{\mathrm{xs}}(\mathbf{R};V,T)\,\mathrm{d}V - S_W(\mathbf{R};V,T)\,\mathrm{d}T \quad (10)$$

$$\overline{\mathbf{f}}_I(\mathbf{R};V,T) \equiv \langle \mathbf{f}_I(\mathbf{r};V)\rangle_{\mathbf{R};V,T} \quad (11)$$

$$\overline{p}_{\mathrm{xs}}(\mathbf{R};V,T) \equiv \langle p_{\mathrm{xs}}(\mathbf{r};V)\rangle_{\mathbf{R};V,T} \quad (12)$$

where $\mathbf{f}_I(\mathbf{r};V)$ is the force on site I and $(\mathrm{d}\mathbf{R}_I)_V = V^{1/3}\mathrm{d}\hat{\mathbf{R}}_I$ indicates changes in the CG configuration at constant volume. While Eq. (**??**) describes variations in an energy, Eq. (10) describes variations in a free energy, including both energetic and entropic contributions.

A few points should be noted:

1. Most importantly, Eq. (10) equates the configuration-, volume-, and temperature-derivatives of the PMF to conditioned averages of the excess forces, excess pressure, and excess entropy of the atomic model. Consequently, the state-point dependence of the PMF, e.g., with respect to temperature or volume change, is determined by

the contributions of the missing atomic degrees of freedom to the conjugate excess thermodynamic quantity, i.e., the excess entropy or excess pressure, respectively. This is the origin of both transferability and representability problems.

2. The third entropic contribution to Eq. (10) is unique to the CG model. Since $S_W \leq 0$, increasing temperature will cause the PMF to increase at each $\mathbf{R}$. In particular, the PMF will vary more rapidly with temperature for CG configurations that correspond to highly structured atomic distributions.

3. Finally, these three contributions are all inter-related via Maxwell-type relations for mixed second derivatives of the PMF. For instance:

$$\left(\partial \bar{\mathbf{f}}_I(\mathbf{R}; V, T) \big/ \partial T\right)_{\mathbf{R},V} = \left(\partial S_W(\mathbf{R}; V, T) \big/ \partial \mathbf{R}_I\right)_{T,V} , \tag{13}$$

which suggests that the temperature-transferability of CG force fields can be maximized by minimizing the configurational dependence of S_W.

2.6 Pressure and the constant NPT ensemble

According to Eq. (**??**), W accounts for the excess, but not the ideal, contributions to the Helmholtz potential from the atoms that have been eliminated from the CG model. Consequently, W does not reproduce the internal pressure and does not provide appropriate Boltzmann weight for sampling different volumes at constant external pressure, $\mathrm{P_{ext}}$. Accordingly, the PMF must be slightly modified in order to model the atomic pressure and the constant NPT ensemble:

$$W_P(\mathbf{R}, V; T) = W(\mathbf{R}; V, T) - (n - N) k_B T \ln\left(V/V_0\right) , \tag{14}$$

where V_0 is an arbitrary reference volume. The second term in Eq. (**??**) accounts for the ideal contributions to the free energy from the missing atomic degrees of freedom. Although

this term does not impact the configuration-distribution at a given V, it ensures that W_P provides the correct Boltzmann weight for each CG microstate, $(\mathbf{R}, V)$:

$$\exp\left[-\beta\left(W_P(\mathbf{R},V;T)+\mathrm{P}_{\mathrm{ext}}V\right)\right]=V_0^{-(n-N)}\int_V \mathrm{d}\mathbf{r}\ \exp[-\beta\left(u(\mathbf{r};V)+\mathrm{P}_{\mathrm{ext}}V\right)]\,\delta\left(\mathbf{M}(\mathbf{r})-\mathbf{R}\right), \tag{15}$$

where $\beta = 1/k_BT$.[36,38] Because

$$-\left(\partial W_P(\mathbf{R},V;T)/\partial V\right)_{\hat{\mathbf{R}};T}=-\left(\partial W(\mathbf{R};V,T)/\partial V\right)_{\hat{\mathbf{R}},T}+(n-N)k_BT/V, \tag{16}$$

according to Eq. (**??**), W_P is the appropriate potential for reproducing the average pressure of the atomic model in each CG microstate,

$$P_{\mathrm{int}}(\mathbf{R},V;T;W_P)=\langle p_{\mathrm{int}}(\mathbf{r};V,T)\rangle_{\mathbf{R};V,T}\ , \tag{17}$$

and each thermodynamic equilibrium state:

$$P_{\mathrm{int}}(V,T;W_P)=p_{\mathrm{int}}(V,T). \tag{18}$$

3 Approximate coarse-graining

The preceding analysis not only clarifies their common origin, but also suggests practical computational methods for resolving representability and transferability challenges.

3.1 Pressure-matching

In practice, CG models commonly employ a relatively simple effective potential that is independent of both temperature and volume, i.e., $U = U_R(\mathbf{R})$.[7,8] The excess pressure of

the CG model is then

$$P^0_{\rm xs}(\mathbf{R};V) \equiv -\left(\partial U_R(\mathbf{R})/\partial V\right)_{\hat{\mathbf{R}}} \quad (19)$$

$$= \frac{1}{3V}\sum_I \mathbf{F}_I(\mathbf{R})\cdot\mathbf{R}_I = \frac{1}{3V}\sum_{(I,J)} F_{2;IJ}(R_{IJ})R_{IJ}, \quad (20)$$

where the last expression sums over each pair (I,J), assuming that each nonbonded interaction is modeled with a pair force function $F_{2;IJ}$ as a function of the pair distance, R_{IJ}.

Leading bottom-up methods parameterize U_R to reproduce atomic structural distributions, such as radial distribution functions (RDFs), at a single thermodynamic state point.[8] While this structure-based approach addresses the configuration-dependence of the PMF, it provides little or no insight into the volume- and temperature-dependence of the PMF. Consequently, there is no reason to expect that the resulting model will accurately describe thermodynamic properties or provide an accurate description at other state points.

In particular, in order for the CG model to accurately describe the pressure of the atomic model, it is necessary that $P^0_{\rm xs}(\mathbf{R};V)$, given by Eq. (20), accurately approximates $-\left(\partial W_P(\mathbf{R},V;T)/\partial V\right)_{\hat{\mathbf{R}};T}$. However, while Eq. (20) assumes that the CG interactions are pair-additive and do not explicitly depend upon the volume, W_P describes many-body interactions that explicitly depend upon the volume. In fact, one expects that contributions to W_P that vary only weakly with (or are independent of) CG configuration provide cohesion that significantly reduces the pressure.[45,48] These contributions are effectively invisible to structure-based methods that focus on reproducing RDFs, which are primarily determined by short-ranged, rapidly varying repulsive potentials.[49] Thus, it is unsurprising that bottom-up CG models tend to dramatically over-estimate the internal pressure.[24,26]

Following Das and Andersen (DA),[38] we have adopted a more general form for the approximate potential in order to model the volume-dependence of the PMF[36]

$$U(\mathbf{R},V) = U_R(\mathbf{R}) + U_V(V). \quad (21)$$

The interaction potential, U_R, is optimized to approximate the configuration dependence of the PMF via standard structure-based methods.[8] Since U_R does not explicitly depend upon V and is (usually) pair-additive, it contributes $P^0_{\rm xs}$ to the pressure according to Eq. (20). Conversely, U_V does not impact the equilibrium configuration distribution of the CG model, but directly contributes to the pressure:

$$P_{\rm xs}(\mathbf{R}, V) = P^0_{\rm xs}(\mathbf{R}; V) + F_V(V), \tag{22}$$

where $F_V = -\mathrm{d}U_V/\mathrm{d}V$ is a "pressure correction." Similarly, because

$$\left(\frac{\partial P_{\rm xs}(\mathbf{R}, V)}{\partial V}\right)_{\hat{\mathbf{R}}} = \left(\frac{\partial P^0_{\rm xs}(\mathbf{R}; V)}{\partial V}\right)_{\hat{\mathbf{R}}} - \frac{\mathrm{d}^2 U_V(V)}{\mathrm{d}V^2}, \tag{23}$$

the second derivative of U_V directly contributes to the (inverse) compressibility. Consequently, U_V can be constructed to accurately model the pressure, the compressibility, and more generally the pressure equation of state for the atomic model. Note that employing a potential that "actively" varies with the density can introduce modifications to the chemical potential, which must be considered to reconcile the virial and compressibility routes for calculating the pressure.[50]

Given a fixed interaction potential, U_R, DA proposed optimizing U_V by minimizing a "pressure-matching" functional:

$$\chi^2_2[U] = \left\langle \left| p_{\rm int}(\mathbf{r}; V, T) - P_{\rm int}(\mathbf{M}(\mathbf{r}), V; T; U) \right|^2 \right\rangle_{PT}, \tag{24}$$

in which the average is evaluated over the constant NPT ensemble for the atomic model. Subsequently, we developed a self-consistent pressure-matching approach that optimizes U_V to quantitatively reproduce the atomic pressure equation of state.[36] This iterative pressure-matching approach corresponds to variationally minimizing a relative entropy[51] with respect

to U_V:[37]

$$S_{\rm rel}[U] = \int {\rm d}V \int_V {\rm d}\mathbf{R}\, p_{RV}(\mathbf{R},V) \ln\left[p_{RV}(\mathbf{R},V)/P_{RV}(\mathbf{R},V;U)\right], \qquad (25)$$

where $p_{RV}(\mathbf{R},V)$ and $P_{RV}(\mathbf{R},V;U)$ are equilibrium distributions for the atomic and CG models, respectively, at constant $\mathrm{P_{ext}}$ and T. It should be noted that W_P minimizes both χ_2^2 and $S_{\rm rel}$. However, given the approximate potential in Eq. (**??**), minimizing $S_{\rm rel}$ ensures that the CG model reproduces the atomic pressure equation of state, while minimizing χ_2^2 does not ensure such consistency.

3.2 Numerical results

Recently, we tested the pressure-matching approach for molecular liquids.[36] Figures 2 and 3 compare the density fluctuations, pressure equations of state, and pressure-volume fluctuations obtained from constant NPT simulations of the OPLS-AA model for heptane[52] and from simulations of several 3-site CG models. The black points in Fig. 3b present a scatter plot of the volume and instantaneous pressure sampled by the OPLS-AA model. The corresponding black curves in Figs. 2 and 3a present the simulated volume fluctuations and pressure equation of state, respectively.

Given these atomic simulations, we employed the multiscale coarse-graining (MS-CG) variational principle[41,53,54] to determine an interaction potential, U_R, for 3-site CG models. This MS-CG potential quite accurately described the structure of liquid heptane, but dramatically overestimated the pressure of the OPLS-AA model. The cyan points in Fig. 3b present a scatter plot of the instantaneous pressure that is generated by applying the MS-CG interaction potential to the configurations sampled by the OPLS-AA model, i.e., $Nk_BT/V + P^0_{\rm xs}(\mathbf{M}(\mathbf{r});V)$. Consequently, constant NPT simulations with the MS-CG interaction potential (without including a pressure correction) overestimated the volume of the OPLS-AA model by more than 10%, as indicated by the blue curve in Fig. 2.

Given this MS-CG interaction potential, we then employed the DA pressure-matching

variational principle[38] to determine a volume-dependent potential, U_V. As indicated by the green curves in Figs. 2 and 3, the resulting DA model much more accurately described the OPLS-AA pressure-volume behavior. Finally, we iteratively refined U_V via self-consistent pressure-matching. The red curves in Figs. 2 and 3 demonstrate that the resulting DN model reproduced the equilibrium density, pressure, and compressibility of the OPLS-AA model with nearly quantitative accuracy.

Figures 2 and 3 also provide instructive comparisons with experiment and with a top-down model. The orange curves present results inferred from experimental measurements of the equilibrium density and compressibility of heptane.[55] The purple curves present results for a top-down model, which Shinoda, Devane, and Klein (SDK) parameterized to reproduce the bulk density and liquid-vapor surface tension, but not the compressibility, of heptane.[56] Because it accurately describes the volume-dependence of the PMF, the bottom-up DN model reproduces experimental measurements of the equilibrium density and compressibility with comparable, if not better, accuracy than the top-down SDK model. Thus, Figs. 2 and 3 demonstrate the promise of bottom-up CG methods for accurately describing both structural and thermodynamic properties.

We have performed self-consistent pressure-matching for 1-, 2-, and 3-site CG heptane models, for 1- and 3-site toluene models, and for 3-site models of heptane-toluene mixtures. In each case, we reproduced the atomic density, compressibility, and pressure equation of state with nearly quantitative accuracy. Interestingly, the optimized pressure correction always dramatically reduced the internal pressure of the CG model, while the ideal kinetic contribution from the "missing atoms" corresponded to a much smaller increase in pressure. As illustrated in Fig. 4a, increasingly large pressure corrections are required with increased coarsening, as the MS-CG interaction potentials systematically lose cohesion, due to the increasingly entropic character of the PMF,[35] as reflected by reduced structure in the site-site RDFs.[37] Figure 4b demonstrates the same correlation between cohesion and pressure correction among MS-CG models for heptane-toluene mixtures of varying composition. Thus,

the pressure correction appears to compensate for the cohesion that is lost in structure-based potentials, as suggested by the classic van der Waals picture.[45,48]

3.3 Transferability for mixtures

The state-point dependence of the PMF limits the transferability of approximate CG potentials. We previously proposed an extended ensemble approach for determining transferable potentials that optimally approximate the PMF across a range of thermodynamic conditions.[39] We have recently combined the extended ensemble and pressure matching approaches to develop predictive CG models for accurately modeling the structure and pressure-volume behavior of heptane-toluene mixtures.[37]

We first employed a global force-matching variational principle to determine a single set of system-independent, i.e., transferable, interaction potentials that accurately approximate the configuration-dependence of the PMF for a range of mixtures. Given this set of transferable interaction potentials, we then performed self-consistent pressure-matching to determine an optimal pressure correction for each mixture. Importantly, this pressure correction can be accurately predicted as a function of the mixture composition. Figure 5 demonstrates that the resulting CG models accurately reproduced the pressure-volume behavior not only for the mixtures employed in the parameterization, but also for two additional mixtures that were not included in the parameterization.

4 Conclusion: van der Waals perspective

In closing, we hope that this work helps clarify the origin and inter-relationship of the representability and transferability problems that plague bottom-up coarse-graining approaches. Both arise as a consequence of thermodynamic information that has been extracted from the atomic configuration space and encoded into the many-body PMF. The key to resolving these problems lies in quantifying this information and then incorporating it into the

calculation of thermodynamic properties and the prediction of transferable potentials. In particular, pressure-matching provides a practical and rigorous way for determining the density dependence of the PMF in order to accurately model the pressure equation of state. Thus, bottom-up approaches can develop predictive, transferable potentials that accurately model the structure, density fluctuations, pressure, and compressibility of atomic models.

More generally, this work suggests a "van der Waals" perspective for bottom-up coarse-graining. From this perspective, current bottom-up approaches provide a powerful means for approximating the configuration-dependence of the PMF, such that the resulting models accurately model atomic structure. At the same time, these approaches do not effectively address the thermodynamic information that determines both the state-point dependence of the PMF and also the missing atomic contribution to thermodynamic properties. Complementary variational principles, such as pressure-matching, may provide an effective means for determining this information, both to predict the transferability of approximate potentials and to model thermodynamic properties.

Acknowledgments

W.G.N. gratefully acknowledges very productive collaborations with M. Scott Shell that contributed to part of the work reviewed herein, as well as very helpful comments on this manuscript from Markus Deserno, Lasse Jensen, Ard Louis, and Christine Peter. This work has been financially supported by the National Science Foundation (NSF Grant Nos. MCB-1053970, CHE-1565631), by the Alfred P. Sloan Foundation, and by a Camille Dreyfus Teacher-Scholar Award. This work was also supported by ACS PRF under Grant No. 52100-ND6. We gratefully acknowledge the Donors of the American Chemical Society Petroleum Research fund for support of this research. This work was partially supported by funding from the Penn State Materials Computation Center. Portions of this research were conducted with Advanced CyberInfrastructure computational resources provided by The Institute for

CyberScience at The Pennsylvania State University (http://ics.psu.edu).

Biographical Information

Nicholas J.H. Dunn was born in 1988 in Portland, ME. In 2011, he earned a B.S. in chemistry from Union College. His research focuses on the thermodynamic properties of CG models and on asphaltene aggregation.

Thomas T. Foley was born in 1989 in Marysville, CA. In 2010, he earned a B.S. in physics from the University of California, Santa Barbara. His research focuses on the role of representation on modeling information and physical systems..

William G. Noid was born in 1978 in Knoxville, TN. In 2000, he earned a B.S. in chemistry from the University of Tennessee, Knoxville. In 2005, he earned a Ph.D. in chemistry from Cornell University. He then spent two years as a postdoctoral research fellow at the University of Utah, Salt Lake City. Since arriving at Penn State in 2007, he has lead a research group that develops multiscale modeling methods and investigates the biophysical properties of disordered proteins.

References

1 Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, *42*, 73–93.

2 Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodriguez-Ropero, F.; van der Vegt, N. F. A. Systematic coarse-graining methods for soft matter simulations - a review. *Soft Matter* **2013**, *9*, 2108–2119.

3 Peter, C.; Kremer, K. Multiscale simulation of soft matter systems. *Faraday Disc.* **2010**, *144*, 9–24.

4 Deserno, M. Mesoscopic Membrane Physics: Concepts, Simulations, and Selected Applications. *Macromol. Rapid Comm.* **2009**, *30*, 752–771.

5 Hyeon, C.; Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2011**, *2*, 487.

6 Müller, M.; Katsov, K.; Shick, M. Biological and synthetic membranes: What can be learned from a coarse-grained description? *Phys. Rep.* **2006**, *434*, 113–176.

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

8 Noid, W. G. Systematic methods for structurally consistent coarse-grained models. *Methods Mol Biol* **2013**, *924*, 487–531.

9 Lyubartsev, A. P.; Laaksonen, A. Osmotic and activity coefficients from effective potentials for hydrated ions. *Phys. Rev. E* **1997**, *55*, 5689–5696.

10 Louis, A. A.; Bolhuis, P. G.; Hansen, J. P.; Meijer, E. J. Can polymer coils be modeled as "soft colloids". *Phys. Rev. Lett.* **2000**, *85*, 2522–5.

11 Vettorel, T.; Meyer, H. Coarse graining of short polyethylene chains for studying polymer crystallization. *J. Chem. Theor. Comp.* **2006**, *2*, 616–629.

12 Ghosh, J.; Faller, R. State point dependence of systematically coarse-grained potentials. *Mol. Sim.* **2007**, *33*, 759–767.

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

14 Krishna, V.; Noid, W. G.; Voth, G. A. The multiscale coarse-graining method. IV. Transferring coarse-grained potentials between temperatures. *J. Chem. Phys.* **2009**, *131*, 024103.

15 Chaimovich, A.; Shell, M. S. Anomalous waterlike behavior in spherically-symmetric water models optimized with the relative entropy. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1901–1915.

16 Farah, K.; Fogarty, A. C.; Böhm, M. C.; Müller-Plathe, F. Temperature dependence of coarse-grained potentials for liquid hexane. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2894–902.

17 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.

18 Izvekov, S. Towards an understanding of many-particle effects in hydrophobic association in methane solutions. *J. Chem. Phys.* **2011**, *134*, 034104.

19 Mirzoev, A.; Lyubartsev, A. P. Effective solvent mediated potentials of Na+ and Cl- ions in aqueous solution: temperature dependence. *Phys. Chem. Chem. Phys.* **2011**, *13*, 5722–5727.

20 Louis, A. A. Beware of density dependent pair potentials. *J. Phys.: Condens. Matter* **2002**, *14*, 9187–206.

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

22 Barker, J.; Henderson, D.; Smith, W. Pair and triplet interactions in argon. *Mol. Phys.* **1969**, *17*, 579–592.

23 van der Hoef, M. A.; Madden, P. A. Three-body dispersion contributions to the thermodynamic properties and effective pair interactions in liquid argon. *J. Chem. Phys.* **1999**, *111*, 1520–1526.

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

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

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

27 Baron, R.; de Vries, A. H.; Hünenberger, P. H.; van Gunsteren, W. F. Configurational Entropies of Lipids in Pure and Mixed Bilayers from Atomic-Level and Coarse-Grained Molecular Dynamics Simulations. *J. Phys. Chem. B* **2006**, *110*, 15602–15614.

28 Baron, R.; Molinero, V. Water-Driven CavityLigand Binding: Comparison of Thermodynamic Signatures from Coarse-Grained and Atomic-Level Simulations. *J. Chem. Theor. Comp.* **2012**, *8*, 3696–3704.

29 Lu, J.; Qiu, Y.; Baron, R.; Molinero, V. Coarse-Graining of TIP4P/2005, TIP4P-Ew, SPC/E, and TIP3P to Monatomic Anisotropic Water Models Using Relative Entropy Minimization. *J. Chem. Theor. Comp.* **2014**, *10*, 4104–4120.

30 D'Adamo, G.; Pelissetto, A.; Pierleoni, C. Predicting the thermodynamics by using state-dependent interactions. *J. Chem. Phys.* **2013**, *138*, 234107.

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

32 Clark, A. J.; McCarty, J.; Lyubimov, I. Y.; Guenza, M. G. Thermodynamic Consistency in Variable-Level Coarse Graining of Polymeric Liquids. *Phys. Rev. Lett.* **2012**, *109*, 168301.

33 McCarty, J.; Clark, A. J.; Lyubimov, I. Y.; Guenza, M. G. Thermodynamic Consistency between Analytic Integral Equation Theory and Coarse-Grained Molecular Dynamics Simulations of Homopolymer Melts. *Macromolecules* **2012**, *45*, 8482–8493.

34 McCarty, J.; Clark, A. J.; Copperman, J.; Guenza, M. G. An analytical coarse-graining method which preserves the free energy, structural correlations, and thermodynamic state of polymer melts from the atomistic to the mesoscale. *J. Chem. Phys.* **2014**, *140*, 204913.

35 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.

36 Dunn, N. J. H.; Noid, W. G. Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *J. Chem. Phys.* **2015**, *143*, 243148.

37 Dunn, N. J. H.; Noid, W. G. Bottom-up coarse-grained models with predictive accuracy and transferability for both structural and thermodynamic properties of heptane-toluene mixtures. *J. Chem. Phys.* **2016**, *144*, 204124.

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

39 Mullinax, J. W.; Noid, W. G. Extended Ensemble approach for deriving transferable Coarse-grained potentials. *J. Chem. Phys.* **2009**, *131*, 104110.

40 Tuckerman., M. E. *Statistical Mechanics: Theory and Molecular Simulation*; Oxford University Press: Oxford, Great Britain, 2013.

41 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.

42 Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.

43 Likos, C. N. Effective interactions in soft condensed matter physics. *Phys. Rep.* **2001**, *348*, 267 – 439.

44 Akkermans, R. L. C.; Briels, W. J. A structure-based coarse-grained model for polymer melts. *J. Chem. Phys.* **2001**, *114*, 1020–1031.

45 Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: San Diego, CA USA, 1990.

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

47 Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chem. Rev.* **2010**, *110*, 1463–1497.

48 Weeks, J. D. Connecting Local Structure to Interface Formation: A Molecular Scale van der Waals Theory of Nonuniform Liquids. *Annu. Rev. Phys. Chem.* **2002**, *53*, 533–562.

49 Andersen, H. C.; Chandler, D.; Weeks, J. D. Roles of Repulsive and Attractive Forces in Liquids : The Equilibrium Theory of Classical Fluids. *Adv. Chem. Phys.* **1976**, *34*, 105.

50 Stillinger, F. H.; Sakai, H.; Torquato, S. Statistical mechanical models with effective potentials: Definitions, applications, and thermodynamic consequences. *J. Chem. Phys.* **2002**, *117*, 288–296.

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

52 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*, 11225–36.

53 Izvekov, S.; Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109*, 2469 – 2473.

54 Izvekov, S.; Voth, G. A. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* **2005**, *123*, 134105.

55 Lide, D. R., Ed. *CRC Handbook of Chemistry and Physics, 90th Edition*, 90th ed.; CRC Press: Ann Arbor, MI USA, 2009.

56 Shinoda, W.; Devane, R.; Klein, M. L. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Mol. Sim.* **2007**, *33*, 27–36.

Figures

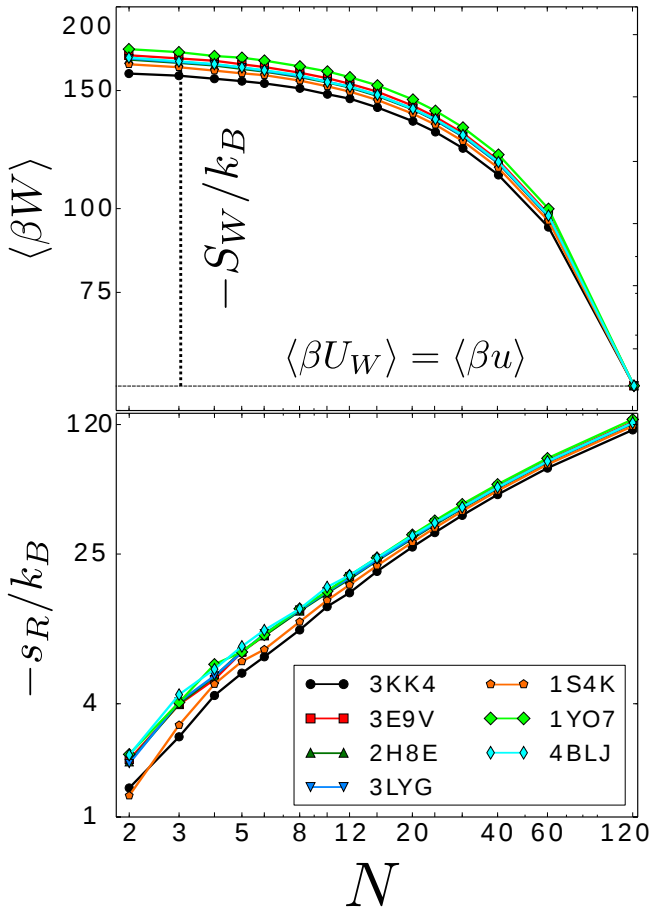


Figure 1: Analysis of the PMF (top) and apparent configurational entropy (bottom) as a function of the number, N, of sites considered. The top panel indicates the energetic (horizontal dotted line) and entropic (vertical dotted line) contributions to the average of the dimensionless PMF (solid line), $\langle \beta W \rangle$, for each protein domain. The bottom panel presents the absolute magnitude of the apparent configurational entropy for each protein domain when viewed at the given resolution. Both panels employ a log x - log y scale. Reproduced from Ref. 35, with the permission of AIP Publishing.

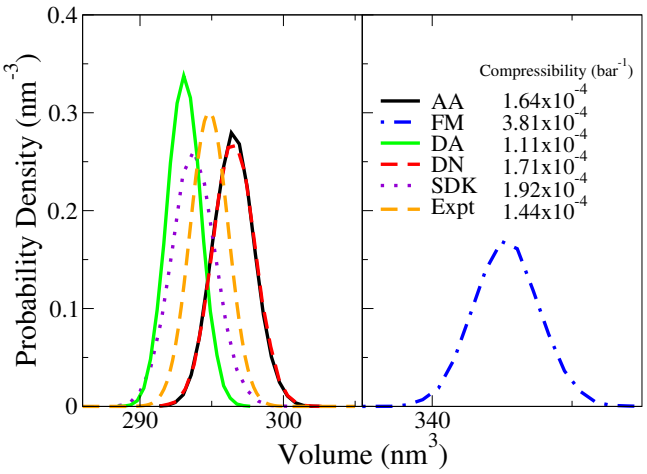


Figure 2: Simulated volume distributions for various heptane models. The solid black curve presents the simulated distribution for the OPLS-AA model. The dashed-dotted blue, solid green, dashed red, and dotted purple curves indicate simulated distributions for the MS-CG, DA, DN, and SDK 3-site models, respectively. The dashed orange curve indicates the normal distribution that is constructed from the experimentally known density and compressibility of heptane. Reproduced from Ref. 36, with the permission of AIP Publishing.

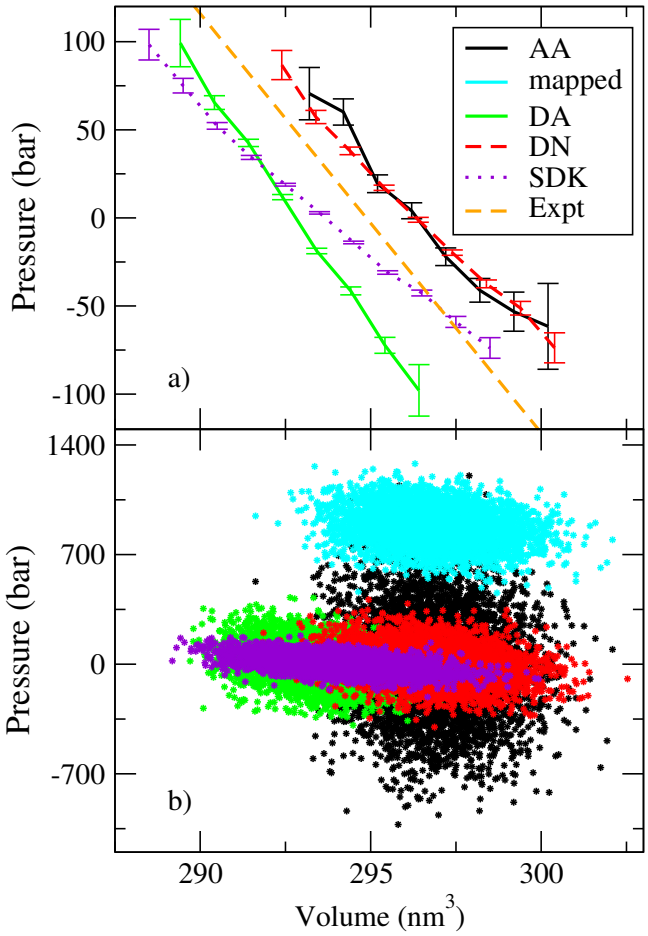


Figure 3: Comparison of the pressure-volume behavior for the AA heptane model and for different 3-site CG heptane models. The black, green, red, and purple curves correspond to the models of Fig. 3. Panel a) presents the equation of state for each model, which is estimated from the mean pressure at each volume in the simulated constant NPT ensemble. The error bars indicate the standard error in the simulated means. The orange curve indicates the equation of state that is determined from the experimentally known density and compressibility of heptane. Panel b) presents a scatter plot of the simulated pressure and volume. The cyan points correspond to the pressure, P^0_{CG}, that is determined by applying the MS-CG potential to the mapped ensemble. Reproduced from Ref. 36, with the permission of AIP Publishing.

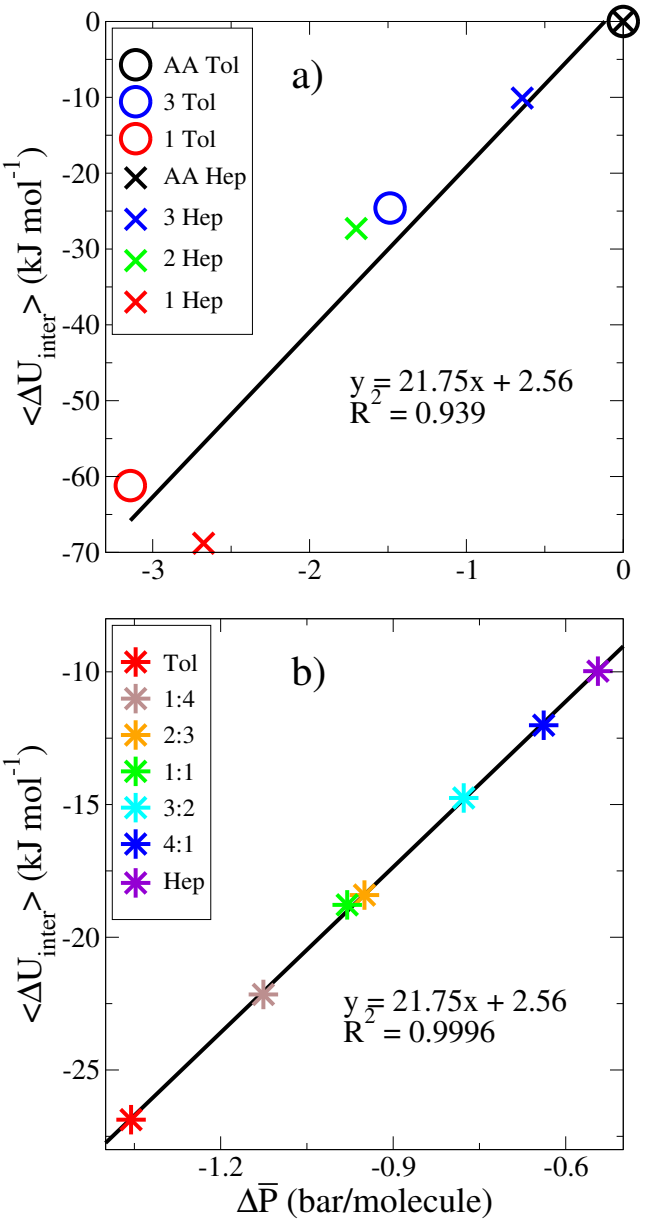


Figure 4: Scatter plot correlating the missing cohesive energy density, $\langle \Delta U_{Inter} \rangle$, with the average pressure correction, $\Delta \overline{P}$, required for various models. Panel a presents results for AA (black), as well as 3-site (blue), 2-site (green), and 1-site (red) CG models for heptane (crosses) and for toluene (circles). Panel b presents results for 3-site CG models of liquid mixtures with varying heptane:toluene mole ratio. The slight differences in the two panels for 3-site models of pure heptane and pure toluene reflect finite size effects.[36–38] Adapted from Refs. 36 and 37, with the permission of AIP Publishing.

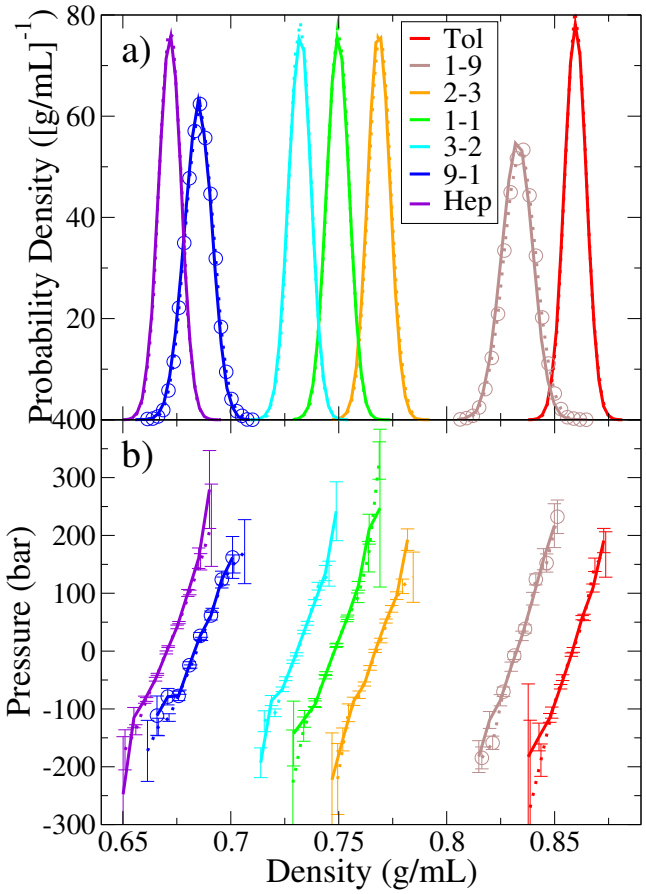


Figure 5: Simulated density distributions (top) and pressure-volume equations of state (bottom) for AA (solid) and 3-site CG (dotted) models for various heptane-toluene mixtures. Pure heptane, pure toluene, as well as the 2:3, 1:1, and 3:2 heptane:toluene mixtures were included in parameterizing the CG potentials. The results for the 1:9 and 9:1 heptane:toluene mixtures were not included in the parameterization and reflect predictions of the transferable model. Adapted from Ref. 37, with the permission of AIP Publishing.