

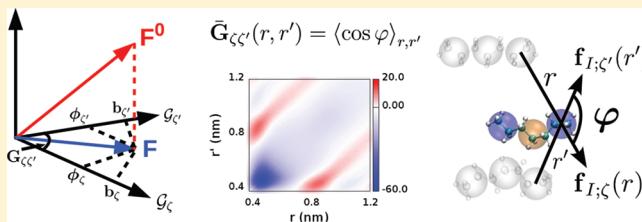
The Role of Many-Body Correlations in Determining Potentials for Coarse-Grained Models of Equilibrium Structure

Joseph F. Rudzinski and William G. Noid*

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, United States

Supporting Information

ABSTRACT: Coarse-grained (CG) models often employ pair potentials that are parametrized to reproduce radial distribution functions (rdf's) determined for an atomistic model. This implies that the CG model must reproduce the corresponding atomistic mean forces. These mean forces include not only a direct contribution from the corresponding interaction but also correlated contributions from the surrounding environment. The many-body correlations that influence this second contribution present significant challenges for accurately reproducing atomistic distribution functions. This work presents a detailed investigation of these many-body correlations and their significance for determining CG potentials while using liquid heptane as a model system. We employ a transparent geometric framework for directly determining CG potentials that has been previously developed within the context of the multiscale coarse-graining and generalized Yvon–Born–Green methods. In this framework, a metric tensor quantifies the relevant many-body correlations and precisely decomposes atomistic mean forces into contributions from specific interactions, which then determine the CG force field. Numerical investigations reveal that this metric tensor reflects both the CG representation and also subtle correlations between molecular geometry and intermolecular packing, but can be largely interpreted in terms of generic considerations. Our calculations demonstrate that contributions from correlated interactions can significantly impact the pair mean force and, thus, also the CG force field. Finally, an eigenvector analysis investigates the importance of these interactions for reproducing atomistic distribution functions.



INTRODUCTION

The complexity and computational expense of atomically detailed models has motivated considerable interest in coarse-grained (CG) models that represent systems in reduced detail.^{1–3} However, the practical utility of these models relies upon potentials that accurately describe the interactions in the low resolution CG model.⁴ Although CG models that are parametrized with thermodynamic data^{5–8} have provided important insight into many processes,^{9,10} these models may not necessarily provide a quantitative description of structural properties.^{11–14}

The correct potential for a CG model that quantitatively reproduces all structural properties of a particular atomistic model (at the resolution of an associated CG mapping) is a configuration-dependent free energy function.¹⁵ This function is defined by the many-body probability distribution for the atomistic model to sample CG configurations (according to this CG mapping).² This function is referred to as a many-body potential of mean force (PMF) because its (negative) gradients define a many-body force that equals the mean atomistic force averaged over all configurations that map to the given CG configuration.¹⁶ Simulations with this many-body mean force (MF) field will quantitatively reproduce the distribution of CG configurations sampled by the atomistic model.¹⁷

In general, though, the many-body PMF cannot be readily calculated, represented, or simulated.¹⁸ Instead, structure-

motivated coarse-graining methods typically approximate the PMF with much simpler molecular mechanics potentials. Each term in this approximate potential models a particular interaction with a function of a single scalar variable (e.g., the distance between a pair of sites or the angle formed by three bonded sites). The atomistic model (along with the CG mapping) defines a target probability distribution for each of these scalar variables. The approximate CG potential is then parametrized to reproduce these atomistic distributions.

Each of these distributions defines a corresponding potential of mean force (or torque¹⁹) along that single degree of freedom. For instance, in the case of nonbonded pair potentials, the distribution of pair distances or, equivalently, the radial distribution function (rdf) determines a pair potential of mean force. The (negative) derivative of this pair potential of mean force defines the pair mean force, which equals the average net force on a particle when a second particle is a given distance away.²⁰ Importantly, if a CG model reproduces the pair mean force of the atomistic model as a function of distance, then the CG model will also reproduce the corresponding atomistic rdf.

Special Issue: Macromolecular Systems Understood through Multiscale and Enhanced Sampling Techniques

Received: January 6, 2012

Revised: May 1, 2012

Published: May 7, 2012

Consequently, the objective of reproducing a set of atomistic rdf's is equivalent to the objective of reproducing a corresponding set of atomistic mean forces. This highlights an important correspondence between iterative approaches that are often expressed in terms of rdf's (e.g., iterative Boltzmann inversion²¹ and the inverse Monte Carlo method²²) and noniterative approaches (e.g., the multiscale coarse-graining^{23,24} and generalized Yvon–Born–Green^{25,26} methods) that can be expressed in terms of the corresponding mean forces.^{27,28} The present work considers structure-motivated coarse-graining from the perspective of reproducing these mean forces.

To briefly introduce the significance of many-body correlations upon the pair mean force, let us consider a monatomic fluid with density, ρ , interacting via pair potentials, $U^{(2)}$. The mean force on particle 1 at position \mathbf{r}_1 when a second particle is located at \mathbf{r}_2 includes both direct and environment-mediated contributions. The direct contribution to this mean force from particle 2 is simply $-\nabla_1 U^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$. However, the environment-mediated contribution reflects many-body correlations and is significantly more complicated. The presence of the two particles impacts the packing of surrounding particles, which are distributed at a position \mathbf{r}_3 with a density $\rho_{3|2}(\mathbf{r}_3|\mathbf{r}_1, \mathbf{r}_2) = \rho g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)/g^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$, where $g^{(2)}$ and $g^{(3)}$ are conventional two- and three-body correlation functions.^{29,30} The Yvon–Born–Green (YBG) integral equation^{29,30} provides a transparent physical picture for decomposing the mean force on particle 1 (i.e., $-\nabla_1 w^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \nabla_1 k_B T \ln g^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$) into direct and environment-mediated contributions:

$$\begin{aligned} -\nabla_1 w^{(2)}(\mathbf{r}_1, \mathbf{r}_2) &= -\nabla_1 U^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \\ &+ \int d\mathbf{r}_3 (-\nabla_1 U^{(2)}(\mathbf{r}_1, \mathbf{r}_3)) \rho_{3|2}(\mathbf{r}_3|\mathbf{r}_1, \mathbf{r}_2) \end{aligned} \quad (1)$$

The net average force from these surrounding particles (i.e., the integral term in eq 1) drives the formation of solvation shells and generates the observed oscillations in the mean force, which correspond to the oscillations in the rdf.

Therefore, even in the simplest (and most common) case that the CG nonbonded potential includes only 2-body interactions, many-body correlations significantly impact the pair mean force and, thus, also the corresponding rdf. Some studies have addressed these effects with more complex, many-body CG potentials, although this can significantly reduce the efficiency of the model.^{31,32} Regardless, any CG model that seeks to accurately reproduce atomistic rdf's must address the many-body correlations that generate the environment-mediated contribution to the mean force.

Various structure-motivated CG approaches differ in their treatment of these many-body correlations. If the density of CG sites is sufficiently low, then the environment-mediated contribution to the mean force vanishes so that the pair mean force in the CG model reflects only the corresponding direct CG force. In this case, direct Boltzmann inversion determines a CG potential that accurately reproduces the atomistic rdf.³³ In some cases, this contribution may be treated via a density expansion.^{34,35} More generally, though, the contributions of many-body correlations to the CG mean forces are treated via iterative methods that perform simulations with trial CG potentials, compare the resulting CG distributions with target atomistic distributions, and then use this information to systematically refine the CG potentials.^{21,22,36–46}

The multiscale coarse-graining (MS-CG) method^{23,24} adopts a considerably different strategy. Each term in the MS-CG

potential defines a corresponding force field basis vector.^{17,47} As illustrated in Figure 1, the MS-CG force field, \mathbf{F} , is defined to

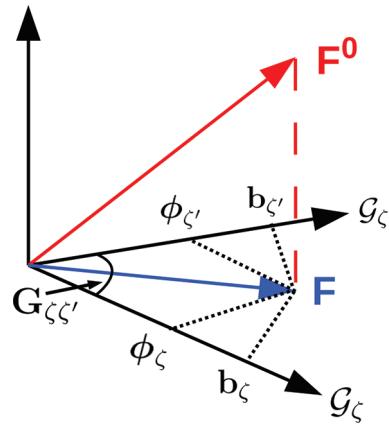


Figure 1. Schematic of the MS-CG procedure. The MS-CG force field, \mathbf{F} , is the projection of the many-body mean force, \mathbf{F}^0 , onto the subspace spanned by the given basis set, including \mathcal{G}_ζ and $\mathcal{G}_{\zeta'}$. The coefficients, $\{\phi_\zeta\}$, of these basis vectors are determined by requiring that \mathbf{F} and \mathbf{F}^0 have the same projections, $\{b_\zeta\}$, onto each basis vector. Since these basis vectors may correspond to correlated interactions, they form a skew coordinate system with a metric tensor $G_{\zeta\zeta'} = \mathcal{G}_\zeta \odot \mathcal{G}_{\zeta'}$.

match the projections of the many-body MF, \mathbf{F}^0 , onto each vector, \mathcal{G}_ζ , in the basis set. Because these basis vectors correspond to correlated interactions, the projections of the MS-CG force field reflect contributions from multiple basis vectors. A metric tensor, $G_{\zeta\zeta'}$, defines the “angle” between basis vectors based upon their correlation. This angle performs a central role in geometrically decomposing the projections of the many-body MF into contributions from individual MS-CG potentials.

Our group has demonstrated that, in the case of nonbonded pair interactions, the corresponding projections of the many-body MF are equivalent to pair mean forces.^{27,28} Moreover, we have demonstrated that the “normal equations” for the MS-CG force field^{47,48} are equivalent to a generalized Yvon–Born–Green (g-YBG) integral equation for molecular mechanics potentials.^{25,26} This g-YBG equation generalizes eq 1 by directly decomposing atomistic pair mean forces into quantitative contributions from each term in the CG potential. This equation provides a computational g-YBG framework for determining the MS-CG “force-matched” forces directly from structures, i.e., without force information.

In summary, because they significantly impact pair mean forces (and thus also the resulting rdf's), all structure-motivated methods must address the many-body correlations that couple interactions in the CG model. The MS-CG metric tensor provides a direct and particularly transparent geometric framework for considering their role in determining potentials for CG models that accurately model molecular structure. Motivated by these considerations, the present work reports a detailed analysis of this metric tensor. In particular, we investigate and characterize the many-body correlations that impact the metric tensor for molecular liquids while using a three-site model of heptane as a model system. Our analysis identifies the key robust features of this metric tensor and precisely elucidates their origin in both structural features of molecular liquids and in the CG representation. We employ the

metric tensor to quantify the contributions of correlated interactions to pair mean forces. Moreover, we perform eigenvector/eigenvalue analysis to identify which correlations are most significant for determining the MS-CG force field and for reproducing atomistic structure. Our calculations and analysis significantly expand upon previous studies that have briefly considered the MS-CG metric tensor.^{48,49} In addition, by focusing on the importance of mean forces in determining potentials for CG models that accurately reproduce atomistic structure, this work complements several previous studies that have compared both formal^{46,50,51} and practical⁵² aspects of force- and structure-based approaches to CG modeling.

The remainder of the paper is organized as follows. Section 2 briefly develops relevant aspects of the MS-CG and g-YBG theory. Section 3 summarizes the key details of our calculations, which are provided in greater detail in the Supporting Information. Section 4 presents a detailed analysis of the metric tensor. Section 5 discusses these results in the context of other recent efforts. Finally, Section 6 summarizes the main conclusions of this work and indicates possible future directions.

THEORY

We briefly summarize relevant aspects of the MS-CG^{17,47,48,53–59} and g-YBG^{25–28} theories and introduce the appropriate notation. This work explicitly considers the canonical ensemble for a system with temperature T and volume V .

The configuration of an atomistic model is defined by the Cartesian coordinates, \mathbf{r} , for n atoms. Similarly, the configuration of a CG model for the same system is defined by the Cartesian coordinates, \mathbf{R} , for N sites. A mapping function, \mathbf{M} , determines a CG configuration as a linear function of the atomic configuration: $\mathbf{R} = \mathbf{M}(\mathbf{r})$. As previously discussed,¹⁷ this definition is sufficiently general for typical CG mappings (e.g., center-of-mass mappings).

The appropriate potential for a CG model that quantitatively reproduces all structural properties of an atomistic model (at the resolution of the CG mapping) is a many-body potential of mean force, $U^0(\mathbf{R})$:

$$U^0(\mathbf{R}) = -k_B T \ln p_R(\mathbf{R}) + \text{const} \quad (2)$$

where $p_R(\mathbf{R})$ is the probability for the atomistic model to sample a configuration that maps to the CG configuration \mathbf{R} . The forces, $\mathbf{F}_i^0(\mathbf{R})$, derived from the many-body PMF define the many-body mean force field and equal the mean force on site I averaged over all atomistic configurations that map to \mathbf{R} .

Because the PMF cannot be readily treated, we consider approximate potentials of a molecular mechanics form,

$$U(\mathbf{R}) = \sum_{\zeta} \sum_{\lambda} U_{\zeta}(\psi_{\zeta}(\{\mathbf{R}\}_{\lambda})) \quad (3)$$

where ζ indicates a particular interaction (e.g., a dihedral angle interaction) and U_{ζ} is the corresponding potential (e.g., a dihedral angle potential) that is a function of a single scalar variable, ψ_{ζ} (e.g., a dihedral angle), that may be expressed as a function of the Cartesian coordinates, $\{\mathbf{R}\}_{\lambda}$, for a set of sites, λ (e.g., the 4 successively bonded sites that form a dihedral angle).⁴⁷ The resulting force on site I may be expressed as

$$\mathbf{F}_I(\mathbf{R}) = \sum_{\zeta} \int dx \phi_{\zeta}(x) \mathcal{G}_{I,\zeta}(\mathbf{R}; x) \quad (4)$$

where

$$\mathcal{G}_{I,\zeta}(\mathbf{R}; x) = \sum_{\lambda} \mathbf{f}_{I,\zeta\lambda}(\mathbf{R}) \delta(\psi_{\zeta\lambda}(\mathbf{R}) - x) \quad (5)$$

$\phi_{\zeta}(x) = -dU_{\zeta}(x)/dx$ is a force function, and $\psi_{\zeta\lambda}(\mathbf{R}) = \psi_{\zeta}(\{\mathbf{R}\}_{\lambda})$. In eq 5, $\mathbf{f}_{I,\zeta\lambda}(\mathbf{R}) = \partial\psi_{\zeta\lambda}(\mathbf{R})/\partial\mathbf{R}_I$ determines the direction of the force on site I from a specific instance, λ , of an interaction of type ζ . We note that the present framework readily generalizes for more complex potentials.⁵¹

The MS-CG approach treats CG force fields as elements in an abstract vector space. Each element in this space defines a set of vector-valued functions that specify a force on each site as a function of the CG configuration. The CG force field defined by eq 4 identifies a particular vector in this space that specifies the set of functions $\mathbf{F} = \{F_1(\mathbf{R}), \dots, F_N(\mathbf{R})\}$. In this framework, the force field can be simply re-expressed:

$$\mathbf{F} = \sum_{\zeta} \int dx \phi_{\zeta}(x) \mathcal{G}_{\zeta}(x) \quad (6)$$

Equation 6 defines a basis set expansion for CG force fields. The set $\{\mathcal{G}_{\zeta}(x)\}$ forms a highly incomplete basis of force field vectors that are determined by the form of the approximate CG potential in eq 3. The corresponding set of force functions, $\{\phi_{\zeta}(x)\}$, act as coefficients that identify a particular element in this vector space. An inner product, \odot , can be defined for any two elements, $\mathbf{F}^{(1)}$ and $\mathbf{F}^{(2)}$, in this vector space: $\mathbf{F}^{(1)} \odot \mathbf{F}^{(2)} = (1/3N) \langle \sum_I \mathbf{F}_I^{(1)}(\mathbf{M}(\mathbf{r})) \cdot \mathbf{F}_I^{(2)}(\mathbf{M}(\mathbf{r})) \rangle$, where the angular brackets denote a canonical average according to the atomistic probability distribution.

Although the many-body MF, \mathbf{F}^0 , is an element in the vector space of CG force fields, in general, it will not be spanned by the basis set included in eq 6. As illustrated in Figure 1, the MS-CG force field is defined by the geometric projection of the many-body MF, \mathbf{F}^0 , onto the vector subspace of force fields spanned by this basis set. Consequently, the many-body MF and MS-CG force field have equal projections along each basis vector.

$$\mathcal{G}_{\zeta}(x) \odot \mathbf{F}^0 = \mathcal{G}_{\zeta}(x) \odot \mathbf{F} \quad (7)$$

The inner product $b_{\zeta}(x) \equiv \mathcal{G}_{\zeta}(x) \odot \mathbf{F}^0$ defines the projection of the many-body MF along the basis vector corresponding to the CG potential $U_{\zeta}(x)$. After expanding \mathbf{F} according to eq 6, we obtain the normal equations^{17,48} for the MS-CG force functions.

$$b_{\zeta}(x) = \sum_{\zeta'} \int dx' G_{\zeta\zeta'}(x, x') \phi_{\zeta'}(x') \quad (8)$$

In eq 8, $G_{\zeta\zeta'}(x, x') \equiv \mathcal{G}_{\zeta}(x) \odot \mathcal{G}_{\zeta'}(x')$ defines a metric tensor that quantifies the “angle” formed between the different basis vectors. Figure 1 and eq 8 demonstrate the underlying simplicity of the MS-CG framework. While the projection $b_{\zeta}(x)$ quantifies an average force along the ψ_{ζ} degree of freedom, the metric tensor decomposes this average force into contributions from each term in the CG potential. Our previous work demonstrated that $b_{\zeta}(x)$ can be expressed in terms of simple structural correlation functions and that the MS-CG equations are equivalent to a g-YBG equation for complex molecular systems.^{25,26} This relation provides the foundation for the g-YBG approach of determining MS-CG potentials directly from structural information.

The metric tensor can be calculated:

$$\begin{aligned} G_{\zeta\zeta'}(x, x') &= \frac{1}{3N} \int d\mathbf{R} p_R(\mathbf{R}) \\ &\times \sum_{\lambda, \lambda'} \left(\sum_I \mathbf{f}_{I, \zeta\lambda}(\mathbf{R}) \cdot \mathbf{f}_{I, \zeta'\lambda'}(\mathbf{R}) \right) \\ &\times \delta(\psi_{\zeta\lambda}(\mathbf{R}) - x) \delta(\psi_{\zeta'\lambda'}(\mathbf{R}) - x') \end{aligned} \quad (9)$$

This average is performed over the CG configuration space, but with the weighting determined by the atomistic model according to p_R . As defined below eq 5, $\mathbf{f}_{I, \zeta\lambda}(\mathbf{R})$ is the direction of the force exerted on site I by the λ instance of the ζ type potential. The associated inner product, that is, $\mathbf{f}_{I, \zeta\lambda}(\mathbf{R}) \cdot \mathbf{f}_{I, \zeta'\lambda'}(\mathbf{R}) = |\mathbf{f}_{I, \zeta\lambda}(\mathbf{R})| |\mathbf{f}_{I, \zeta'\lambda'}(\mathbf{R})| \cos \varphi$, defines an angle φ formed by the forces exerted on the site I by the two particular interactions, $\zeta\lambda$ and $\zeta'\lambda'$. Consequently, $G_{\zeta\zeta'}$ significantly differs from a conventional 2-dimensional probability distribution and may assume positive or negative values. In particular, $G_{\zeta\zeta'}(x, x') \neq 0$ only if (1) there exist configurations, \mathbf{R} , for which $\psi_{\zeta\lambda}(\mathbf{R}) = x$ and $\psi_{\zeta'\lambda'}(\mathbf{R}) = x'$ for some λ and λ' ; (2) the corresponding $\zeta\lambda$ and $\zeta'\lambda'$ interactions exert forces on a shared site, I ; and (3) the corresponding average of $\sum_I \mathbf{f}_{I, \zeta\lambda}(\mathbf{R}) \cdot \mathbf{f}_{I, \zeta'\lambda'}(\mathbf{R})$ does not vanish. Therefore, $G_{\zeta\zeta'}$ may vanish even if the ζ and ζ' interactions are not statistically independent.

By separating out the $\lambda = \lambda'$ term in eq 9, $G_{\zeta\zeta'}(x, x')$ can be decomposed into direct, $\bar{g}_\zeta(x)$, and indirect, environment-mediated, $\bar{G}_{\zeta\zeta'}(x, x')$, contributions: $G_{\zeta\zeta'}(x, x') = \bar{g}_\zeta(x) \delta_{\zeta\zeta'} \times \delta(x - x') + \bar{G}_{\zeta\zeta'}(x, x')$.²⁶ The projections of the many-body MF are then decomposed into direct and correlated contributions:

$$b_\zeta(x) = \bar{g}_\zeta(x) \phi_\zeta(x) + \sum_{\zeta'} \int dx' \bar{G}_{\zeta\zeta'}(x, x') \phi_{\zeta'}(x') \quad (10)$$

In this decomposition,^{26,28} $\bar{g}_\zeta(x)$ is a correlation function of a single variable that is closely related to, for example, conventional rdf's, and $\bar{G}_{\zeta\zeta'}(x, x')$ describes the correlation between the ζ and ζ' interactions and includes only many-body correlations. In analogy to eq 1, the first term in eq 10 identifies the direct contribution of the $\phi_\zeta(x)$ interaction to the projection, $b_\zeta(x)$. As illustrated in Figure 1, $\bar{G}_{\zeta\zeta'}(x, x')$ weights the contributions from the ζ' interaction to $b_\zeta(x)$. We emphasize that $\bar{G}_{\zeta\zeta'}$ differs from $G_{\zeta\zeta'}$ only in that the first sum in eq 9 is restricted to $\lambda \neq \lambda'$. When $\zeta \neq \zeta'$ or $x \neq x'$, $G_{\zeta\zeta'}(x, x') = \bar{G}_{\zeta\zeta'}(x, x')$.

Importantly, this analysis applies for complex potentials with, for example, angle and torsional potentials. However, to make this analysis more concrete and also because many of the following results consider pair nonbonded interactions, we present explicit results for a CG potential that include only a single type of pair potential, $U^{(2)}$. In this case, $U(\mathbf{R}) = \sum_\lambda U^{(2)}(\psi_\lambda(\mathbf{R}))$, λ identifies a particular pair of sites $\{I, J\}$, $\psi_\lambda(\mathbf{R}) = |\mathbf{R}_I - \mathbf{R}_J|$ is the distance between the pair, and $\mathbf{f}_{I, \lambda}(\mathbf{R})$ is a unit vector along $\mathbf{R}_I - \mathbf{R}_J$. Then, $\bar{g}(r) = cr^2 g(r)$ in terms of the rdf, $g(r)$, and the constant $c = (4\pi/3)N/V$. Moreover, $b(r) = k_B T c r^2 dg(r)/dr = -cr^2 g(r) w'(r)$ in terms of the pair mean force, $-w'(r)$, which is defined by the pair potential of mean force,^{20,30} $w(r) = -k_B T \ln g(r)$. The many-body contribution to the metric tensor then simplifies to a sum over triples:

$$\bar{G}(x, x') = \frac{1}{3N} \left\langle \sum_{\lambda}^{\text{triples}} \cos \varphi_\lambda \delta(r_{\lambda J} - x) \delta(r_{\lambda K} - x') \right\rangle \quad (11)$$

where λ identifies the “central” site of a triple; $\{\lambda, J, K\}$, $r_{\lambda J}$ and $r_{\lambda K}$ are the distances from the central site to J and K , respectively; and φ_λ is the angle formed by the triple. Because the number of particles, J , that are a distance, x , away from a central particle λ scales as x^2 , in this simple case, $\bar{G}(x, x')$ should scale with $(xx')^2$. The following calculations explicitly employ this scaling when considering many-body correlations that correspond to nonbonded pair interactions.

More generally, if we consider a particular nonbonded pair potential, ζ , within a more complex CG potential and apply the preceding analysis, eq 10 may be re-expressed as

$$-w'_\zeta(r) = \phi_\zeta(r) + \sum_{\zeta'} \int dx' \frac{1}{c_\zeta g_\zeta(r) r^2} \bar{G}_{\zeta\zeta'}(r, x') \phi_{\zeta'}(x') \quad (12)$$

where c_ζ is a constant, $g_\zeta(r)$ is the atomistic rdf, and $w'_\zeta(r)$ is the atomistic pair potential of mean force for the ζ nonbonded pair interaction. In complete analogy with eq 1, this statement of the g-YBG equation decomposes the pair mean force into a direct force between the pair and correlated contributions from each term in the CG potential. The force functions, $\phi_\zeta(x)$, that satisfy this decomposition of the atomistic pair mean force, $-w'_\zeta(r)$, then determine the MS-CG force field. This relation further clarifies the central role of the metric tensor in the MS-CG method. In particular, if $\bar{G}_{\zeta\zeta'} = 0$, then the MS-CG method reduces to direct Boltzmann inversion.

METHODS

The present section briefly outlines essential details of the following calculations. The Supporting Information provides additional details.

Simulation Details. All molecular dynamics (MD) simulations were performed using the Gromacs 4.5.3 simulation suite⁶⁰ according to standard procedures.^{61–65} MD simulations were performed for three classes of systems: (1) atomistic models of heptane and of water; (2) one-, two-, and three-site MS-CG models of heptane; and (3) model fluids of mono-, di-, and triatomic molecules with geometries similar to the one-, two-, and three-site MS-CG heptane models. The present section briefly outlines the key details of these simulations, which are described in much greater detail in the Supporting Information.

The atomistic heptane simulations considered 267 molecules and modeled all interactions with the OPLS-AA force field.⁶⁶ The heptane model was first equilibrated in the NPT ensemble at 1 bar and 298 K to determine an equilibrium volume of $V = (4.08 \text{ nm})^3$, which agrees within 3% of the experimentally measured density.⁶⁷ The system was then simulated in the NVT ensemble at this volume to determine canonical correlation functions for the following analysis. The Supporting Information also reports details and results for a 20 ns simulation of 216 SPC/E⁶⁸ molecules in the NVT ensemble at 298 K and $V = (1.86 \text{ nm})^3$.

The following subsection describes calculations of the MS-CG force field for each CG heptane representation. An initial configuration for each CG model was obtained by mapping a configuration from all-atom simulations. This configuration was energy-minimized and simulated for 24 ns in the NVT ensemble at 298 K, with the first 4 ns serving for equilibration.

To identify generic features of $\bar{G}_{\zeta\zeta'}$, we simulated model fluids of mono-, di-, and triatomic molecules. In each case, all nonbonded interactions were modeled with identical Lennard-

Jones pair potentials that generated a stable liquid phase. Each system was simulated in the NVT ensemble at 298 K after equilibration in the constant NPT ensemble at 298 K and 1 bar pressure to determine the equilibrium volume.

Mapping. The atomically detailed simulations of heptane were mapped to one-, two-, and three-site CG representations, as shown in Figure 2. The one-site mapping defined a single

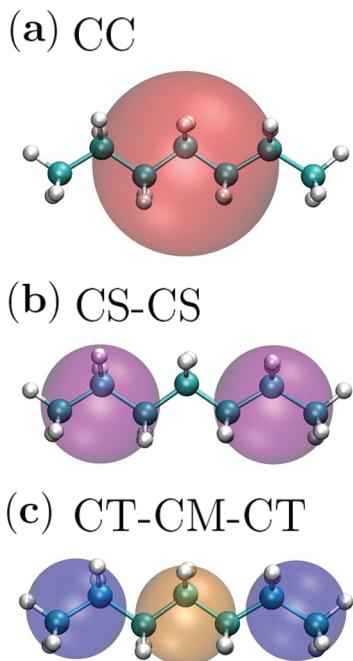


Figure 2. CG representations of heptane. (a) The one-site model defines one CC site (red) for each molecule. (b) The two-site model defines two CS sites (purple) for the two terminal $\text{CH}_3\text{CH}_2\text{CH}_2$ groups. (c) The three-site model defines two CT sites (blue) for the terminal CH_2CH_3 groups and a CM site (orange) for the central $\text{CH}_2\text{CH}_2\text{CH}_2$ group. The coordinates of each site are defined by the center of mass for the associated atomic group.

(CC) site at the molecular center of mass. The two-site mapping defined two equivalent (CS) sites at the centers of mass of the two terminal $\text{CH}_2\text{CH}_2\text{CH}_3$ groups. The three-site mapping defined two equivalent (CT, blue) sites at the centers of mass for the two terminal CH_2CH_3 groups and a single (CM, orange) site at the center of mass for the middle $\text{CH}_2\text{CH}_2\text{CH}_2$ group. Molecular graphics in Figure 2 and elsewhere were rendered with VMD.⁶⁹

Force Field Calculations. For each CG representation of heptane, the MS-CG potential was calculated from a discrete version of the normal equations presented in eq 8. As appropriate, the intramolecular potentials included bond and angle terms. All intermolecular interactions were modeled with central pair potentials. Each potential function was represented by a discrete set of basis functions of a single variable.⁴⁷ The CG force field is then expanded in a corresponding discrete basis set: $\mathbf{F} = \sum_D \phi_D \mathbf{G}_D$, where D identifies a particular basis function used to represent a particular potential function, $U_\zeta(x)$, \mathbf{G}_D is the resulting force field basis vector, and ϕ_D is the coefficient of that vector, which also determines the corresponding force function, $\phi_\zeta(x)$. These basis vectors determine discrete representations of $G_{\zeta\zeta}(x, x')$ and $b_\zeta(x)$, that is, $G_{DD'}$ and b_D . $G_{DD'}$ was calculated from configurations sampled by all-atom MD simulations; b_D was calculated from

atomistic forces according to the MS-CG approach^{47,48,55} and from structural correlation functions according to the g-YBG approach.^{25–28} Both methods quantitatively agreed, as expected. These correlation functions then determine a discrete set of normal equations for the MS-CG force field parameters.

$$\sum_{D'} G_{DD'} \phi_{D'} = b_D \quad (13)$$

These equations are explicitly derived in the Supporting Information and have been previously discussed.^{25,28,47,48,55}

In most calculations, the potentials were represented on a grid, that is, with piecewise constant basis functions.⁴⁸ The bond, angle, and nonbonded pair potentials were represented with grid spacings of 0.0005 nm, 0.5 deg, and 0.005 nm, respectively. In this basis set, $G_{DD'}$ and b_D correspond to representing $G_{\zeta\zeta}(x, x')$ and $b_\zeta(x)$, respectively, on a set of discrete grid points. To investigate the sensitivity of simulated structural properties to the form of the CG potential, the CG force field was also calculated using Lennard-Jones type functions to represent each nonbonded pair potential, that is, $U_\zeta^{n-m}(x) = \phi_{\zeta 1}/x^n + \phi_{\zeta 2}/x^m$, where $\phi_{\zeta 1}$ and $\phi_{\zeta 2}$ may be either positive or negative. We report results for $n - m = 12 - 6$ and 8–4, but similar results were obtained for $n - m = 11 - 6$, 10–6, 10–4, 9–6, and 8–6.

The normal equations (eq 13) were solved via LU decomposition after applying right preconditioning.⁷⁰ The calculated force functions were smoothed with a running average over three consecutive grid points and then integrated to determine corresponding potentials. These potentials were interpolated and employed to simulate each CG model. Pair, three-body, and many-body correlation functions were calculated from these trajectories and compared with those calculated from the mapped, all-atom trajectories. The Supporting Information demonstrates that the MS-CG models quantitatively reproduced the structure of the atomistic OPLS heptane model.

Molecular Interpretation of $\bar{G}_{\zeta\zeta}$. As defined above eq 10, the MS-CG metric tensor $G_{\zeta\zeta}(x, x')$ can be decomposed into a correlation function of a single variable, $\bar{g}_\zeta(x)$, and a many-body correlation function, $\bar{G}_{\zeta\zeta}(x, x')$, that describes the coupling between the ζ and ζ' interactions. As noted above, $\bar{G}_{\zeta\zeta} = G_{\zeta\zeta}$, except for the case that $\zeta = \zeta'$ and $x = x'$. To characterize the relevant many-body correlations, we calculated $\bar{G}_{\zeta\zeta'}$ for each CG mapping shown in Figure 2 while using configurations sampled from atomistic simulations.

Equation 9 demonstrates that $\bar{G}_{\zeta\zeta'}$ differs from conventional many-body correlation functions in two ways: (1) $\bar{G}_{\zeta\zeta'}$ can adopt either positive or negative values based upon the direction of forces from the ζ and ζ' interactions; and (2) $\bar{G}_{\zeta\zeta'}$ reflects only correlations between interactions that exert forces on a shared particle. To better understand $\bar{G}_{\zeta\zeta'}$ and, in particular, to understand this directional effect, we compared $\bar{G}_{\zeta\zeta}(x, x')$ to a more conventional many-body correlation function,

$$P_{\zeta\zeta'}(x, x') = \frac{1}{3N} \int d\mathbf{R} p_R(\mathbf{R}) \sum_{\lambda \neq \lambda'}^* \delta(\psi_{\zeta\lambda}(\mathbf{R}) - x) \times \delta(\psi_{\zeta'\lambda'}(\mathbf{R}) - x') \quad (14)$$

where λ and λ' identify particular instances of the ζ and ζ' interactions, respectively, and the star indicates that the sum is restricted to the set of distinct interactions, $\zeta\lambda$ and $\zeta'\lambda'$, that

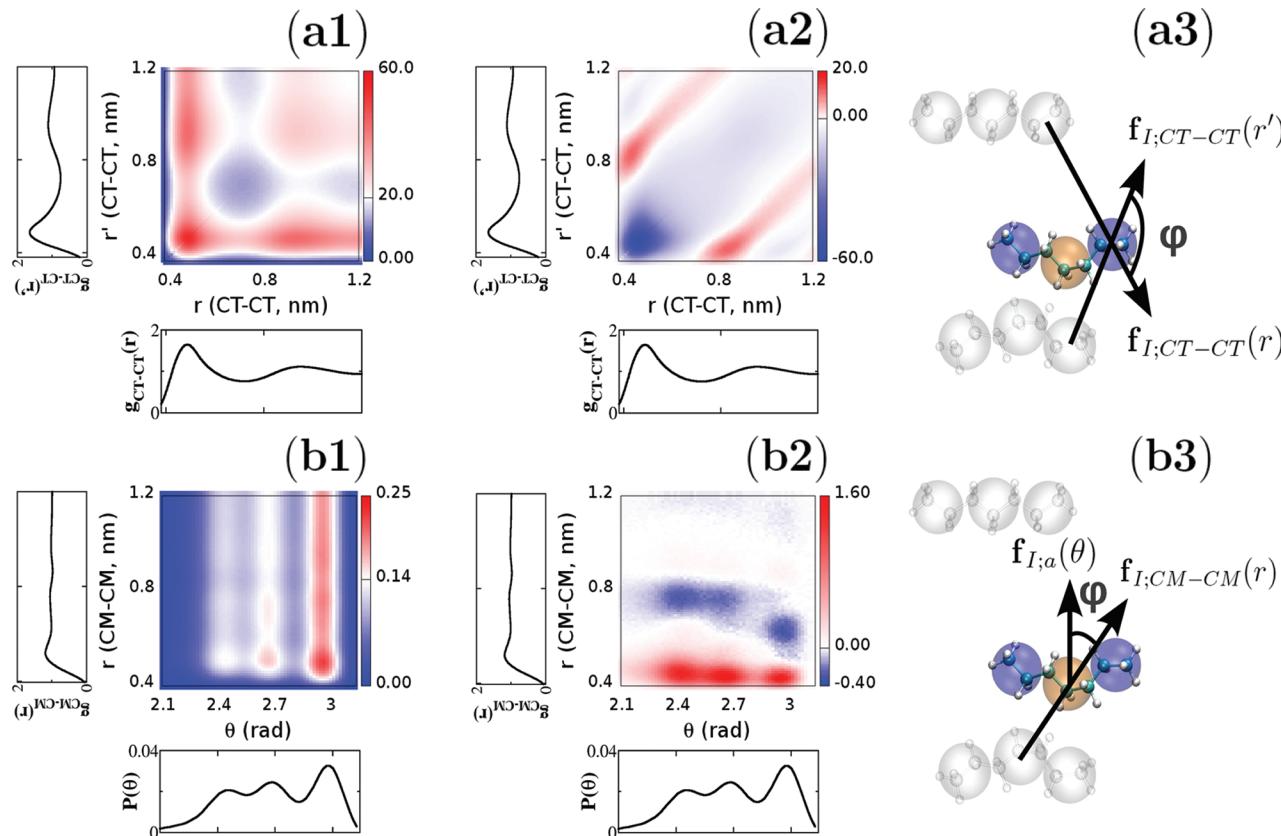


Figure 3. Many-body correlations in the OPLS-AA heptane model. Row a describes correlations between CT–CT and CT–CT pair nonbonded interactions. Row b describes correlations between the CM–CM pair nonbonded interaction and the intramolecular angle interaction. Columns 1 and 2 quantify these correlations with intensity plots of $P_{\zeta\zeta}$ and $\bar{G}_{\zeta\zeta}$, respectively. Column 3 illustrates the angle, φ , formed by the force vectors, $\mathbf{f}_{I;\zeta\zeta}$, that contribute to $\bar{G}_{\zeta\zeta}$ in each case. For clarity of presentation, $P_{\zeta\zeta}$ and $\bar{G}_{\zeta\zeta}$ are rescaled by factors of 10^2 and 10^3 , respectively, in this figure and in Figure 4.

exert forces on a shared particle. With this definition, both $\bar{G}_{\zeta\zeta}$ and $P_{\zeta\zeta}$ include contributions from the same set of interactions; however, $P_{\zeta\zeta}$ weights each contribution with equal (positive) weight, whereas $\bar{G}_{\zeta\zeta}$ weights each contribution on the basis of the geometry of the interaction.

As discussed above, if ζ corresponds to a nonbonded pair interaction, then both $\bar{G}_{\zeta\zeta}(x, x')$ and $P_{\zeta\zeta}(x, x')$ scale according to x^2 . In analogy to the definition of the rdf^{20,30} and to focus on the local many-body correlations that influence pair mean forces and CG force fields, the following section presents $\bar{G}_{\zeta\zeta}(x, x')$ and $P_{\zeta\zeta}(x, x')$ after rescaling both according to $(r_\zeta(x) r_\zeta(x'))^{-2}$, where $r_\zeta(x) = x$ when ζ corresponds to a nonbonded interaction and 1, otherwise. These correlation functions are presented as intensity plots using gnuplot 4.4.0.⁷¹

Eigenvalue Analysis of $G_{\zeta\zeta}$. To investigate the role of many-body correlations in determining the CG force field, we performed eigenvalue/eigenvector analysis of $G_{\zeta\zeta}$. Upon determining the eigenvalues, $\{\lambda_i\}$, and associated eigenvectors, $\{\mathbf{v}_i\}$, of G_{DD} , eq 13 may be expressed as

$$\mathbf{b} = \sum_i \lambda_i \phi_{\lambda_i} \mathbf{v}_i \quad (15)$$

where \mathbf{b} is a vector with elements b_D and ϕ_{λ_i} is the component along the eigenvector \mathbf{v}_i of an analogous vector of force field coefficients, $\boldsymbol{\phi}$. These eigenvectors identify correlated forces that contribute to the projections of the CG force field.

The following section considers eigenvalues and eigenvectors calculated after rescaling $G_{\zeta\zeta}(x, x')$ by $1/(r_\zeta^2(x) r_\zeta^2(x'))$. We

employ this scaling for several reasons: (1) This rescaling corresponds to a simple redefinition of the basis vectors and, equivalently, to natural left and right preconditioning of $G_{\zeta\zeta}(x, x')$ to solve eq 13; (2) this rescaling corresponds to the representation employed in visualizing the structural correlation function $\bar{G}_{\zeta\zeta}(x, x')$; and (3) if rescaling is not applied, then the calculated eigenvalue spectrum is dominated by long-ranged nonbonded interactions, simply due to the x^2 scaling of nonbonded pairs, as discussed above.

We employed LAPack (Linear Algebra PACKage)⁷² to calculate the eigenvalues and eigenvectors of $G_{\zeta\zeta}(x, x')/r_\zeta^2(x) r_\zeta^2(x')$ for the three-site heptane representation. Given this set of calculated eigenvalue/eigenvector pairs, we defined a participation fraction, $n_{i,\zeta}$, to quantify the significance of the ζ interaction to eigenvector i . For each eigenvector i , we calculated $n_{i,\zeta}$ by adding the absolute values of the eigenvector elements for the basis vectors, $\{\mathcal{G}_\zeta(x)\}$, associated with the potential $U_\zeta(x)$. We then normalized this quantity for each eigenvector, i , by the sum of the absolute values for all the elements in the eigenvector.

We calculated an analogous quantity to determine the contributions to each eigenvector from short-, medium-, and long-ranged nonbonded interactions. In this calculation, interactions were considered short-ranged if they corresponded to distances smaller than the first minima in the corresponding calculated CG potential. This distance is approximately 0.5 nm for each interaction. Interactions were considered medium-ranged if they corresponded to distances between the first

potential minimum and the first local maximum in the force function at approximately 0.8 nm. Interactions corresponding to greater distances were considered long-ranged.

RESULTS

Structure-motivated CG approaches must address the effects of many-body correlations when determining potentials for reproducing atomistic structure. The MS-CG metric tensor, $G_{\zeta\zeta}(x, x')$, provides a transparent, geometric framework both for characterizing these correlations and also for quantifying their impact upon the CG potential. The following calculations provide a detailed analysis of $G_{\zeta\zeta}$ in the context of determining MS-CG potentials for the one-, two-, and three-site models of liquid heptane that are shown in Figure 2. We first precisely relate robust aspects of $G_{\zeta\zeta}$ to specific structural features of the atomistic model and to the CG representation, respectively. We then demonstrate the role of $G_{\zeta\zeta}$ in decomposing atomistic pair mean forces into contributions from the various terms in the approximate CG force field. Finally, we employ eigenvalue/eigenvector analysis to investigate the relationship between the metric tensor, the calculated MS-CG force field, and the equilibrium structure of the resulting CG model. Although it is not the focus of the present work, the Supporting Information demonstrates that the one-, two-, and three-site MS-CG heptane models each accurately reproduce the structure of the atomistic model.

Molecular Interpretation of $\bar{G}_{\zeta\zeta}$. The function $\bar{G}_{\zeta\zeta}$ quantifies the contributions of many-body correlations to the MS-CG metric tensor. To relate $\bar{G}_{\zeta\zeta}$ to specific structural features of molecular liquids, Figure 3 compares $\bar{G}_{\zeta\zeta}$ with the more conventional many-body correlation function $P_{\zeta\zeta}$, which is defined in eq 14. Both correlation functions were computed using configurations that were sampled from atomistic simulations of the OPLS heptane model and then mapped to the three-site heptane representation shown in Figure 2c. As discussed above, in the cases that ζ (or ζ') corresponds to a potential of a distance, r_ζ , the correlation functions $P_{\zeta\zeta}$ and $\bar{G}_{\zeta\zeta}$ have been rescaled by r_ζ^2 to highlight the relevant local structural features.

Figure 3a considers the case that both ζ and ζ' correspond to the CT–CT nonbonded pair potential. Figure 3a1 presents an intensity map of $P_{\zeta\zeta}(r, r')$, which quantifies the probability of observing a set, $\{\lambda, J, K\}$, of three CT sites in which sites J and K (shadowed in Figure 3a3) are distances r and r' , respectively, from the central λ site (blue in Figure 3a3). In this intensity plot (and in Figure 3b1), blue, red, and white identify regions of zero, high, and intermediate probability, respectively. Figure 3a1 demonstrates regions of high and low probability in $P_{\zeta\zeta}$ that correspond to solvation shells in the CT–CT rdf.

Figure 3a2 presents the corresponding block of $\bar{G}_{\zeta\zeta}(r, r')$. In Figures 3 and 4, negative, positive, and zero values of $\bar{G}_{\zeta\zeta}(r, r')$ are indicated by blue, red, and white regions, respectively. By definition, both $P_{\zeta\zeta}(r, r')$ and $\bar{G}_{\zeta\zeta}(r, r')$ include contributions from the same sets of triples. However, although each triple contributes equal (positive) weight to $P_{\zeta\zeta}(r, r')$, the triple, $\{\lambda, J, K\}$, contributes $\cos \varphi$ to $\bar{G}_{\zeta\zeta}(r, r')$, with φ being the angle between the nonbonded CT–CT force vectors, as illustrated in Figure 3a3. In particular, eq 11 demonstrates that $\bar{G}_{\zeta\zeta}(r, r')$ vanishes if for given r and r' , all angles, φ , are equally sampled. The nonvanishing regions in Figure 3a2 indicate preferred angles, φ , for particular distances, r and r' .

The most striking feature of Figure 3a2 is a large negative peak located at $r \approx r' \approx 0.5$ nm. This feature corresponds to

triples for which CT sites J and K are both in the first solvation shell of the central site, λ . Supporting Figure S10 demonstrates that the excluded volume of sites J and K prevents their overlap, excluding configurations for which $\cos \varphi \geq 0.75$ while also promoting configurations for which $\cos \varphi = -1$. The resulting slight preference for the sites to pack on opposite sides of λ and for the force vectors to form an obtuse angle (i.e., $\cos \varphi < 0$) generates the observed negative peak. This steric effect generates a negative band for the diagonal $r \approx r'$ region but becomes decreasingly significant with increasing r .

Figure 3a2 also demonstrates positive bands at $r' \approx r \pm \sigma$, where σ is the site diameter, which is estimated by the first peak in the corresponding rdf. These positive bands also reflect the solvation shell structure. Supporting Figure S10 demonstrates that, for configurations with $r' \approx r \pm \sigma$, the sites J and K are slightly more likely to be on the same side of the central site, λ . Consequently, the angle φ is slightly more likely to be acute, and the resulting average will be positive. This effect is most pronounced when r' is slightly less than $r + \sigma$, since the triplet configuration is more likely to be slightly staggered than perfectly colinear.

Columns 1 and 2 of Figure 3b present a similar comparison of $P_{\zeta\zeta}$ and $\bar{G}_{\zeta\zeta}$ for the case that ψ_ζ corresponds to the intramolecular CT–CM–CT angle, θ , and $\psi_{\zeta'}$ corresponds to the CM–CM pair distance, r . Figure 3b1 demonstrates three distinct bands in $P_{\zeta\zeta}(\theta, r)$ that correspond to heptane rotamer states, with the largest band corresponding to an all-trans configuration. Within these bands, regions of high and low intensities correspond to the maxima and minima in the CM–CM rdf.

Figure 3b2 presents the corresponding block of $\bar{G}_{\zeta\zeta}(\theta, r)$. Figure 3b3 demonstrates the direction of the force vectors from ζ and ζ' interactions that act on a shared CM site. In the configurations mapped from the atomistic simulation, the force vector from the angle potential approximately bisects the intramolecular angle, θ . As in Figure 3a, the nonbonded CM–CM pair force is directed along the vector connecting the two sites.

Although the three bands observed in $P_{\zeta\zeta}(\theta, r)$ remain intact, the peaks in these bands now demonstrate alternating signs in $\bar{G}_{\zeta\zeta}(\theta, r)$. Supporting Figure S12 demonstrates that, as CM sites approach one another, the flanking CT sites must project away. As illustrated in Figure 3b3, the organization of CT sites away from the approaching CM site results in an acute angle (i.e., $\cos \varphi > 0$) between the associated force vectors and generates a positive peak in $\bar{G}_{\zeta\zeta}$. This steric effect is most pronounced for molecules with a relatively small intramolecular angle, $\theta \approx 2.4$ rad. Consequently, in comparison to $P_{\zeta\zeta}(\theta, r)$, the band of $\bar{G}_{\zeta\zeta}(\theta, r)$ that corresponds to small intramolecular angles, $\theta \approx 2.4$ rad, grows in magnitude relative to the band for $\theta \approx 3.0$ rad. The alternating signs of the peaks in each band result from the packing of molecules into solvation shells on opposite sides of a central molecule, as described in the context of Figure 3a2. Supporting Figure S13 suggests that the shift in this negative peak for $\theta \approx 3.0$ rad results from subtle packing effects.

In summary, Figure 3 demonstrates that $\bar{G}_{\zeta\zeta}$ reflects relatively subtle aspects of intermolecular packing and its coupling to intramolecular configuration; however, these features result from generic properties of soft condensed-phase systems, e.g., steric interactions, solvation shell packing, and molecular geometry. To investigate the generality of these features and also the sensitivity of the metric tensor to the CG

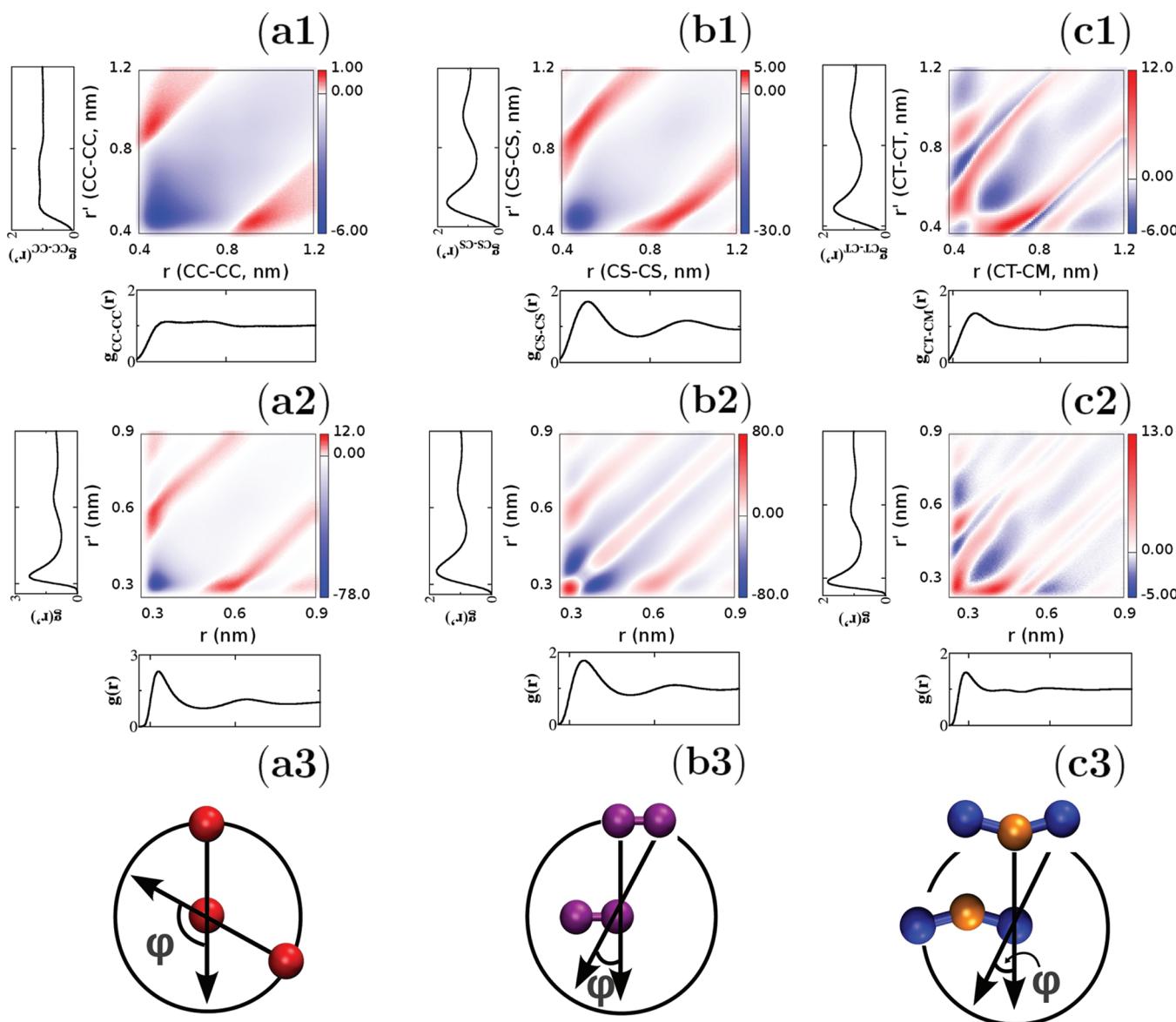


Figure 4. Intensity plots of $\bar{G}_{\zeta\zeta}$ for nonbonded interactions in heptane and associated model fluids. Columns 1, 2, and 3 correspond to one-, two-, and three-site models. Rows 1 and 2 present correlation functions calculated using configurations sampled for the OPLS heptane model and for corresponding model fluids, respectively. The interactions in row 1 are identified by site types labeled in Figure 2. Finally, row 3 illustrates the angle, φ , that contributes to $\bar{G}_{\zeta\zeta}$.

representation, we calculated $\bar{G}_{\zeta\zeta}$ for model mono-, di-, and triatomic fluids and compared these results with calculations for the OPLS heptane simulations, while using one-, two-, and three-site mappings (Figure 2).

Figure 4 presents these calculations. In the first row, columns a, b, and c present results obtained by mapping atomistic heptane simulations to one-, two-, and three-site representations, respectively. In the second row, columns a, b, and c present calculations for simulations of model mono-, di-, and triatomic fluids, respectively. The third row illustrates the relevant interactions for each calculation. In each column, ζ and ζ' identify nonbonded pair potentials for corresponding site pairs in the two models. The Methods section and Supporting Information provide details of these calculations.

Column a of Figure 4 compares calculations of $\bar{G}_{\zeta\zeta}$ for a one-site (CC) representation of atomistic heptane simulations (Figure 4a1) and for a model monatomic fluid (Figure 4a2). The size and asymmetry of heptane result in much broader and

more diffuse bands than are observed for the spherically symmetric monatomic fluid. Nevertheless, Figure 4a1 and a2 demonstrate the same generic features observed in Figure 3a2.

Column b of Figure 4 compares calculations of $\bar{G}_{\zeta\zeta}$ for a two-site (CS-CS) representation of atomistic heptane simulations (Figure 4b1) and for a model diatomic fluid (Figure 4b2). Figure 4a1 and b1 demonstrate that $\bar{G}_{\zeta\zeta}$ has similar structure for the one- and two-site representations of heptane. However, Figure 4b2 reveals a new positive (red) feature along the diagonal of $\bar{G}_{\zeta\zeta}$ for the model diatomic fluid. As indicated in Figure 4b3, if the intramolecular bond length is sufficiently short, then the presence of a single site at a distance r away significantly increases the probability of a second site at a similar distance. The force vectors from these bonded sites form an acute angle (i.e., $\cos \varphi > 0$) and result in this positive feature for $r \approx r'$. The width of this feature corresponds to the intramolecular bond length and its magnitude decays with increasing r . Supporting Figure S16 demonstrates that this

positive band vanishes if the bond length of the model diatomic becomes sufficiently large relative to the corresponding site diameters. This is, indeed, the case for the two-site heptane mapping, in which both the bond length and the site diameter are ~ 0.5 nm.

Finally, column c of Figure 4 compares calculations of $\bar{G}_{\zeta\zeta}$ for a three-site (CT–CM–CT) representation of atomistic heptane simulations (Figure 4c1) and for a model triatomic fluid (Figure 4c2). In this case, ζ and ζ' correspond to the CT–CM and CT–CT pair nonbonded interactions. The model triatomic corresponding to Figure 4c2 was designed so that the ratio of the bond length, d , to the site diameter, σ , is very similar for the two models. As a result, the features of $\bar{G}_{\zeta\zeta}(r, r')$ in Figure 4c1 and c2 are remarkably similar. In both cases, the bond length is sufficiently large that the positive diagonal feature observed in Figure 4b2 has disappeared. Instead, the region corresponding to $r' \approx r$ reflects simple excluded volume effects similar to Figure 3a2. In addition, Figure 4c1 and c2 demonstrate a positive band at $r' \approx r \pm \sigma$ corresponding to the second solvation shell, as in Figure 3a2. However, Figure 4c1 and c2 also demonstrate a new positive feature at $r' \approx r \pm d$ (with $d < \sigma$) corresponding to the forces from the two CT sites that flank each central CM site and that exert forces in the same direction, as suggested by Figure 4c3.

Decomposition of Pair Mean Forces. Figure 1 and eq 8 demonstrate that $G_{\zeta\zeta}$ decomposes projections, $b_\zeta(x)$, of the many-body MF into contributions from each force field basis vector. In the case that ζ corresponds to a bonded or nonbonded pair potential, this projection is closely related to the atomistic pair mean force, $-w'_\zeta(r)$, according to eq 12. The decomposition of this atomistic mean force then defines the MS-CG force functions, $\{\phi_\zeta(r)\}$.

Columns 1, 2, and 3 of Figure 5 demonstrate this decomposition for the CT–CT, CM–CM, and CT–CM pair nonbonded interactions, respectively. In each panel, the solid

black curve presents the pair mean force and the solid red curve presents the contribution from the corresponding direct force. In each case, the marked difference between the mean and direct forces quantifies the significance of many-body correlations for determining additional contributions to the mean force. These differences are particularly significant at short distances near the first peak of each rdf. At these distances, the mean forces are much more attractive than the corresponding direct forces. (Note that the corresponding atomistic rdf attains local maxima or minima when the corresponding pair mean force vanishes and, in each case, achieves its first maxima near 0.5 nm.)

The dashed green and dashed blue curves in Figure 5a present the contributions to the nonbonded pair mean forces from intramolecular angle and bond forces, respectively. These bonded interactions contribute significantly, especially at short distances, and demonstrate the importance of the many-body correlations in Figure 3b for coupling intra- and intermolecular interactions.

Figure 5b quantifies the contributions to each nonbonded pair mean force (solid black) from the corresponding direct force (solid red) and from correlated CT–CT (dashed green), CM–CM (dashed cyan), and CT–CM (dashed magenta) nonbonded forces. In particular, Figure 5b1 demonstrates that, when a pair of CT sites is separated by 0.8 nm, the direct force between the pair is attractive, but the average force on each site is actually repulsive. This repulsive mean force results largely from correlated forces due to other CT sites in the first solvation shell. (See Supporting Information Figure S18.) Figure 5b1 demonstrates that correlated forces from these CT sites also significantly reduce the CT–CT mean force at short distances. Similar short-ranged effects are observed for each pair interaction.

Figure 5b demonstrates that correlated CT–CM and CT–CT interactions significantly impact both CT–CT and the CM–CM mean forces (Figure 5b2 and b3, respectively), suggesting that these interactions are particularly important for reproducing the pair structure of the system. In contrast, the CM–CM interaction contributes relatively little to the CT–CM mean force. Finally, although the CM–CM and CT–CT pair interactions are statistically correlated (Supporting Information Figure S19), the corresponding block of the metric tensor vanishes because these interactions never exert forces on the same particle. Consequently, the CM–CM, and CT–CT interactions do not contribute to the CT–CT (Figure 5a2) and CM–CM (Figure 5b2) mean forces, respectively.

Eigenvalue Analysis of $G_{\zeta\zeta}$. The MS-CG metric tensor, $G_{\zeta\zeta}(x, x')$, defines a linear transformation from CG force functions, $\phi_\zeta(x)$, to projections, $b_\zeta(x)$, of the CG force field onto basis vectors, $\mathcal{G}_\zeta(x)$. Figure 6 presents eigenvalue/eigenvector analysis of the metric tensor to identify and quantify the correlated interactions that contribute to matching projections of the many-body MF. Each eigenvector identifies a set of correlated interactions that contribute to b_ζ , and the corresponding eigenvalue weights this contribution. For reasons discussed in the Methods section, we calculated the 977 eigenvalue/eigenvector pairs for a discrete matrix representation of $G_{\zeta\zeta}(x, x')/(r_\zeta^2(x) r_\zeta^2(x'))$, rather than for $G_{\zeta\zeta}(x, x')$. Figure 6c2 presents the resulting eigenvalue spectrum, which is bounded above by one (as is the eigenvalue spectrum for $G_{\zeta\zeta}$) and demonstrates two inflection points near $i \approx 350$ and $i \approx 800$. Consequently, eigenvalues (and the associated eigenvectors) will be characterized as small, medium,

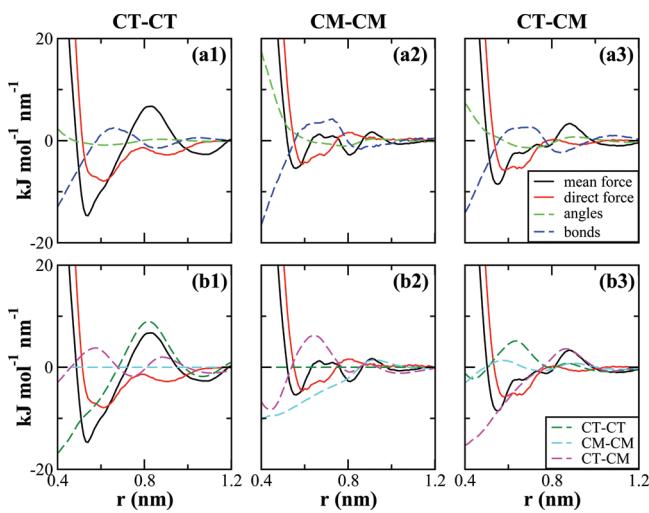


Figure 5. Decomposition of mean forces for the CT–CT (left), CM–CM (center), and CT–CM (right) nonbonded pair interactions. The solid black and red curves represent the calculated mean, $-w'_\zeta(r)$, and direct, $\phi_\zeta(r)$, forces, respectively. The dashed light green and blue curves present correlated contributions to the mean force from intramolecular angle and bond interactions, respectively. The dashed dark green, cyan, and magenta curves present contributions to the mean force from correlated CT–CT, CM–CM, and CT–CM pair nonbonded interactions, respectively.

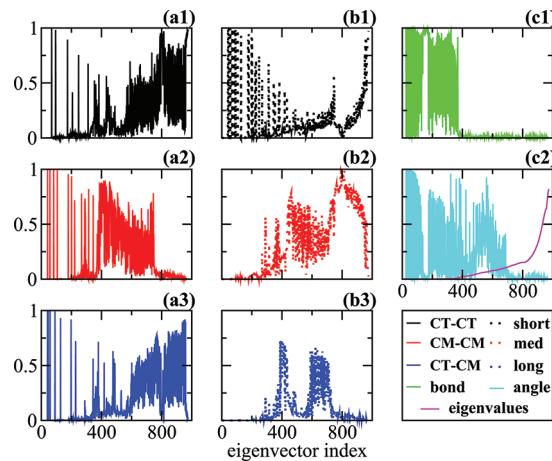


Figure 6. Analysis of the eigenvectors and eigenvalues for the normalized metric tensor calculated using the three-site heptane mapping. In column a, rows 1, 2, and 3 present the participation fractions for CT–CT, CM–CM, and CT–CM nonbonded pair interactions, respectively. In column b, rows 1, 2, and 3 present corresponding quantities for short-, medium-, and long-ranged nonbonded interactions, respectively. In column c, rows 1 and 2 present corresponding quantities for intramolecular bond and angle interactions, respectively. Finally, Figure 5c2 also presents the eigenvalue spectrum in magenta.

or large on the basis of whether $350 > i$, $800 > i > 350$, or $i > 800$, respectively.

As described in the Methods section, we calculated a participation fraction, $n_{i\zeta}$, to quantify the contribution from each interaction ζ to eigenvector i . The participation fraction, $n_{i\zeta}$, vanishes if eigenvector i does not include any contribution from the ζ interaction and is normalized so that $n_{i\zeta} = 1$ if

eigenvector i includes only contributions from U_ζ . Intermediate values of $n_{i\zeta}$ identify eigenvectors that reflect correlated contributions to b_ζ .

Figure 6a1, a2, and a3 present $n_{i\zeta}$ as a function of eigenvector index i for ζ corresponding to the CT–CT, CM–CM, and CT–CM nonbonded interactions, respectively. Figure 6b1, b2, and b3 present an analogous participation fraction that quantifies the contribution to eigenvector i from short- ($r \lesssim 0.5$ nm), medium- (0.5 nm $\lesssim r \lesssim 0.8$ nm), and long-ranged ($r \gtrsim 0.8$ nm) interactions, respectively. Figure 6c1 and c2 present this participation fraction, $n_{i\zeta}$, for ζ corresponding to intramolecular bond and angle interactions, respectively.

Figure 6a demonstrates that the largest eigenvectors, corresponding to the most significant contributions to b_ζ , reflect CT–CT and CT–CM nonbonded interactions. Medium eigenvectors reflect CM–CM nonbonded interactions that are coupled with CT–CT and CT–CM interactions, as well as with intramolecular angle interactions. The smallest eigenvectors correspond to contributions from intramolecular bond and angle interactions. Figure 6b demonstrates that the very largest eigenvectors of the normalized matrix reflect short-ranged interactions, whereas the next largest eigenvectors reflect correlations between short- and medium-ranged interactions. Eigenvectors corresponding to $350 < i < 800$ primarily reflect coupling between medium- and long-ranged interactions. Although Figure 6b1 suggests that very small eigenvectors also reflect short-ranged interactions, these small eigenvectors correspond to very short-ranged interactions that are only rarely sampled.

The inverse of the metric tensor determines a linear transformation from projections, $b_\zeta(x)$, of the many-body MF field to MS-CG force functions, $\phi_\zeta(x)$. While Figure 5 employed the metric tensor to decompose atomistic pair mean forces into specific contributions from MS-CG force

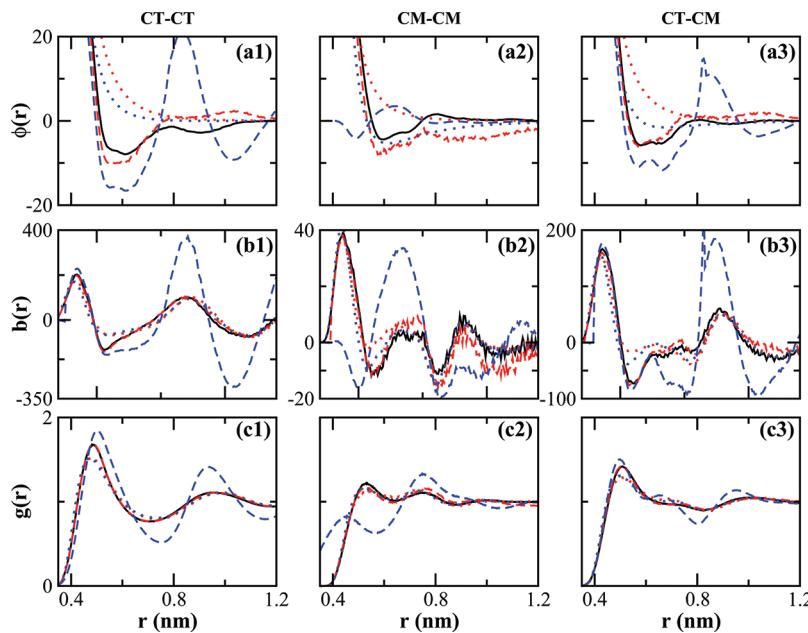


Figure 7. Eigenvector analysis of the MS-CG nonbonded pair forces (row 1), corresponding force projections (row 2), and resulting rdf's (row 3). Rows 1 and 2 are presented in units of $\text{kJ mol}^{-1} \text{nm}^{-1}$. Columns 1, 2, and 3 present results for the CT–CT, CM–CM, and CT–CM nonbonded pair forces, respectively. In each case, the solid black curves present results for the MS-CG force field when nonbonded forces were represented on a grid, whereas the dashed red and dashed blue curves present results after eliminating the contributions of 350 and 800 eigenvectors, respectively, from these nonbonded forces. The dotted red and dotted blue curves present results for the MS-CG force field when using 12–6 and 8–4 Lennard-Jones type functions to represent the nonbonded forces.

functions, the present eigenanalysis decomposes these MS-CG forces into contributions from individual eigenvectors. In particular, the largest eigenvalues correspond to those interactions that are most significant for reproducing projections, $b_\zeta(x)$, of the many-body MF and that are, consequently, most robustly determined by the MS-CG method. The smallest eigenvalues correspond to interactions that are least significant for reproducing $b_\zeta(x)$ and that are least well determined. By systematically eliminating the contributions of specific eigenvectors from the MS-CG force field, we investigate the sensitivity of the MS-CG model to various features of the CG potential and also to many-body correlations present in the atomistic model.

Figure 7 presents this analysis. In each case, the solid black curves correspond to the MS-CG force field calculated by numerically solving eq 13. The dashed red and dashed blue curves correspond to results after eliminating the contributions to the MS-CG nonbonded forces from the smallest 350 and smallest 800 eigenvectors, respectively. The left, center, and right columns of Figure 7 correspond to results for the CT–CT, CM–CM, and CT–CM nonbonded interactions, respectively.

Row a of Figure 7 quantifies the sensitivity of the MS-CG nonbonded forces to these eigenvectors. The solid black curves present the MS-CG force functions, which correspond to the solid red curves in Figure 5. The dashed red curves in Figure 7a demonstrate that the smallest 350 eigenvectors are relatively insignificant for representing the CT–CM force function, but are more significant for the CT–CT and CM–CM forces. As expected from Figure 6, the smallest eigenvectors contribute minimally to repulsive short-ranged forces, but are important for representing longer-ranged forces. The dashed blue curves demonstrate that only 177 eigenvectors are necessary to accurately represent the short-ranged repulsion between CT–CT and CT–CM pairs, as suggested by Figure 6. In contrast, the 177 largest eigenvectors are not sufficient to represent even the short-ranged component of the CM–CM force.

Figure 7b presents projections of these CG force fields onto force field basis vectors for the nonbonded pair interactions. In each case, these projections were calculated by numerically evaluating the left-hand side expression of eq 8. The solid black curves present projections of the MS-CG force field, which quantitatively agree with the corresponding projections of the many-body MF, that is, $F^0 \odot G_\zeta(x)$ in eq 7. As described above, these projections are closely related to the atomistic pair mean forces (black curves) in Figure 5. In particular, the first peaks in the atomistic rdf's correspond to $b_\zeta(r) = 0$ at $r \approx 0.5$ nm.

The dashed red curves in Figure 7b demonstrate that these projections of the many-body MF can be quite accurately reproduced after eliminating the contributions of 350 eigenvectors from the CG nonbonded force functions. (In calculating Figure 7b and c, the contributions of these eigenvectors were not eliminated from the intramolecular force field.) In particular, $b_\zeta(r)$ is reproduced with near-quantitative accuracy for the CT–CT and CT–CM interactions. In contrast, the dashed blue curves demonstrate that the 177 largest eigenvectors are not sufficient to accurately reproduce $b_\zeta(r)$, although the resulting force field reasonably estimates $b_\zeta(r)$ for the CT–CT and CT–CM interactions at short distances.

Finally, the third row of Figure 7 presents rdf's calculated from simulations with each CG force field. The black curves

present rdf's from simulations of the MS-CG model. Supporting Information Figure S7 demonstrates that these rdf's agree quantitatively with corresponding atomistic rdf's. The dashed red (blue) curves present rdf's from simulations with the CG force field after eliminating the smallest 350 (800) eigenvectors from the nonbonded force functions. The dashed red curves demonstrate that these contributions from the 350 smallest eigenvectors are not necessary to reproduce the atomistic mean forces. In fact, the resulting CG simulations quantitatively reproduce the atomistic rdf's. However, as expected, the dashed blue curves demonstrate that the nonbonded interactions defined by only 177 eigenvectors are insufficient to reproduce the atomistic CM–CM rdf. Nevertheless, this very reduced CG force field is sufficient to qualitatively reproduce the atomistic rdf's for the CT–CT and CT–CM pairs. Rows b and c suggest that, in the case of liquid heptane, the accuracy of the CG model in reproducing the projections of the many-body MF provides a predictive indicator of the accuracy with which simulations of the CG model will reproduce atomistic rdf's.

We have also performed similar eigenvector analysis of $\bar{G}_{\zeta\zeta}(x, x')/r_\zeta^2(x)r_\zeta^2(x')$. Supporting Information Figure S21 presents this correlation function after eliminating the smallest 350 and 800 eigenvectors. The correlation functions corresponding to the CT–CT and CT–CM interactions are only slightly impacted by the removing these eigenvectors. On the other hand, the correlation functions related to the CM–CM interaction are more significantly impacted. After eliminating the contributions from 800 eigenvectors, these CM–CM correlation functions are not even qualitatively described.

These results suggest that projections of the CG force field and also the rdf's generated by CG simulations are quite insensitive to many features of the CG force field. In practice, it may be possible to accurately reproduce atomistic rdf's with relatively simple potentials. To test this hypothesis, we recalculated the MS-CG force field, while using simple Lennard-Jones type $U_\zeta^{n-m}(x)$ functions to represent nonbonded pair potentials. In this case, the three nonbonded interactions were determined by a total of six parameters instead of the 527 parameters required in the tabulated representation. (As before, intramolecular interactions were tabulated on a grid.) The dotted red and dotted blue curves in Figure 7 present the results for $n - m = 12-6$ and $8-4$, respectively.

Figure 7a demonstrates that these Lennard-Jones type nonbonded force functions are almost completely repulsive and differ significantly from the force functions calculated for the more flexible discrete delta basis set (solid black curves), although the sets of forces agree quite well in the repulsive hard-sphere regions. Nevertheless, despite these differences, Figure 7b demonstrates that the $n - m$ force functions quite accurately reproduce the projections, b_ζ , of the many-body MF (solid black curve). These results demonstrate that force field projections calculated with $G_{\zeta\zeta}$ diminish differences between CG force fields, which is consistent with the earlier observation that the corresponding eigenvalues are all less than 1. Finally, the dotted curves in Figure 7c demonstrate that MD simulations with these $n - m$ force functions also reproduce the atomistic rdf's with reasonable accuracy. These results provide further evidence that the rdf's calculated from CG MD simulations are similarly insensitive to many details of the CG force fields.

■ DISCUSSION

Structure-motivated coarse-graining approaches typically approximate the many-body PMF with molecular mechanics potentials. Each term in this potential models a particular interaction with a function of a single scalar variable. These approximate potentials are parametrized to reproduce atomistic distribution functions corresponding to each of these individual variables (e.g., rdf's). This objective is equivalent to requiring that the CG model reproduce the atomistic mean force²⁹ (or torque¹⁹) for each interaction. However, the mean forces generated by the CG model include not only a direct contribution from the corresponding interaction but also correlated contributions from the environment.⁴⁸ The many-body correlations that generate these correlated contributions represent a significant challenge for developing CG models that accurately reproduce atomistic structure. An improved understanding of the impact of many-body correlations for determining CG potentials and for the structural accuracy of the resulting models should lead to optimized CG mappings,⁷³ improved approximations to the many-body PMF, and perhaps even meaningful a priori estimates of errors in CG models.⁴⁵

Various approaches treat these correlations in different ways. Direct Boltzmann inversion assumes that nonbonded pair mean forces in the CG model reflect only the corresponding direct force. This approach completely neglects the effects of correlated interactions and, depending upon the density of CG sites, may provide limited accuracy for modeling the structure of complex condensed phases. Methods such as Iterative Boltzmann Inversion (IBI) and Inverse Monte Carlo require multiple simulations to assess the significance of many-body correlations in the CG model and to systematically refine the CG potential. Although these methods typically focus upon rdf's, Soper⁷⁴ demonstrated that IBI updates the CG force functions with the difference between the atomistic and CG pair mean forces. At the same time, the force-based MS-CG method directly determines CG forces to reproduce atomistic pair mean forces via equations that are equivalent to a g-YBG equation.^{25,26} Consequently, mean forces provide an intriguing connection between iterative structure-based and direct force-based methods for determining CG potentials.

As illustrated in Figure 1, the MS-CG and g-YBG methods address the effects of many-body correlations via a geometric framework that is approximate, but direct and transparent. This approach defines projections, $b_\zeta(x)$, of the many-body MF onto each vector, $\mathcal{G}_\zeta(x)$, in an incomplete basis set that is determined by the approximate CG force field.^{17,47} The metric tensor, $G_{\zeta\zeta}(x, x')$, decomposes each projection $b_\zeta(x)$ into (1) a direct contribution from the corresponding force function, $\phi_\zeta(x)$; and (2) correlated contributions from the other CG force functions, $\{\phi_\zeta(x')\}$.

The present calculations demonstrate that $G_{\zeta\zeta}$ reflects not only local packing properties in the atomistic model but also the CG representation. In particular, $G_{\zeta\zeta}$ is quite sensitive to relatively subtle effects, such as the impact of rotameric states upon the asymmetry of surrounding solvation shells. This coupling between inter- and intramolecular interactions suggests that the common practice of determining intramolecular potentials from direct Boltzmann inversion may adversely impact the intermolecular structure of a CG model. In the case of heptane, $G_{\zeta\zeta}$ can be interpreted by simple models with similar geometry (Figure 4). In more complex systems, though, strong specific interactions may bias particular

configurations and generate distinct features in $G_{\zeta\zeta}$. For instance, calculations for the SPC/E water model⁶⁸ demonstrate that hydrogen bonding generates features that cannot be attributed to generic packing properties. (Supporting Information Figure S26.) Consequently, we anticipate that $G_{\zeta\zeta}$ may be particularly useful, not only for determining accurate CG potentials, but also for identifying interactions that are essential for stabilizing complex molecular structures.

In addition to providing efficient computational models, structure-motivated CG approaches that approximate the many-body PMF can also provide quantitative insight into the physical forces that underly particular phenomena. The many-body PMF is a configuration-dependent, free-energy function that reflects both energetic and entropic contributions.^{15–17} Consequently, several insightful studies have employed the individual terms in approximate CG potentials to quantify the thermodynamic forces underlying, for example, hydrophobic self-assembly⁷⁵ and cellulose dissolution⁷⁶ and to decompose these interactions into energetic and entropic components.^{59,75} However, this analysis is somewhat complicated, since it is only in combination that these effective potentials rigorously approximate a free energy function.

The present work suggests a somewhat more precise analysis. The pair potential of mean force for a particular interaction describes the reversible work (i.e., free energy change) along that reaction coordinate.²⁹ Equation 12 decomposes the corresponding pair mean force into specific contributions from each term in the CG potential. In the specific case of a three-site model for heptane, Figure 5 demonstrates that correlated interactions significantly impact the mean force. In particular, at short distances near the first peak of the rdf, these correlated forces dramatically reduce the mean force relative to the corresponding direct force. In addition, this analysis highlighted the effects of nearby neighbors upon the mean force for particles that are separated by greater distances. We note that this analysis depends on the interactions considered in the decomposition and, thus, must be interpreted with caution. Nevertheless, we anticipate that this framework may prove useful in quantifying specific contributions to potentials of mean force for complex molecular systems.

To further investigate the role of the many-body correlations in determining the MS-CG force field, we performed an eigenvector analysis after rescaling $G_{\zeta\zeta}$ according to Figures 3 and 4. The largest eigenvalues of the rescaled matrix, which contribute most to projections of the CG force field, reflect short- and medium-ranged components of the CT–CT and CT–CM interactions. In contrast, most of the smallest eigenvalues reflect intramolecular and long-ranged interactions.

We employed these eigenvectors to represent the MS-CG force field and, in particular, to eliminate the contributions of small eigenvectors from the nonbonded force functions. These calculations further demonstrated that projections, $\mathcal{G}_\zeta(x) \odot F$, onto force field basis vectors, $\mathcal{G}_\zeta(x)$, are relatively insensitive to many features of the CG force field, F . In particular, additional calculations with Lennard-Jones type potentials demonstrated that markedly different pair potentials generated quantitatively similar force field projections. These results are consistent with the observation that the eigenvalue spectrum of $G_{\zeta\zeta}$ is bounded above by 1. Consequently, when determining these projections, $G_{\zeta\zeta}$ contracts the difference between CG force fields. This may be a relatively general property of $G_{\zeta\zeta}$, although further analysis is required to assess this supposition.

We also performed MD simulations with these different CG force fields to assess their accuracy in reproducing atomistic rdf's. Simulations with the MS-CG force field (defined by a flexible discrete delta basis) quantitatively reproduced the structure of the OPLS-AA heptane model; however, additional MD simulations demonstrated that the CG rdf's are remarkably insensitive to the contributions of the 350 smallest eigenvectors to the MS-CG pair potentials. MD simulations with significantly different Lennard-Jones type pair potentials (determined by MS-CG calculations) also reproduced the atomistic rdf's with near quantitative accuracy. These results are consistent with many previous studies, indicating that simulated rdf's are sensitive to short-ranged repulsive potentials, but relatively insensitive to other features of the CG potential.^{21,77}

Our results also indicate that, at least for heptane, the accuracy of a CG force field in reproducing the projections, $b_\zeta(x) \equiv G_\zeta(x) \odot F^0$, of the many body MF, F^0 , provides a reasonable indicator for the accuracy of the CG force field in reproducing the atomistic rdf's. This observation is consistent with the remarkable success of the MS-CG method in modeling atomistic structure.^{23,24,78–83} However, this observation is also somewhat surprising. The CG rdf's are generated by MD simulations and reflect a complicated and nonlinear relationship between the CG potentials and the resulting rdf's. In contrast, according to eq 8, the projections of the CG force field are linearly related to the CG potential.

As noted above, the MS-CG method determines a CG force field that quantitatively matches projections of the many-body MF, that is, $G_\zeta(x) \odot F = G_\zeta(x) \odot F^0$. In the case of nonbonded interactions, these projections have obvious physical significance because $G_\zeta(x) \odot F^0$ corresponds to an atomistic mean force.^{25,28} If simulations of the CG model reproduce these atomistic mean forces, then the CG model will also reproduce the corresponding rdf's. Nevertheless, the MS-CG potentials are not guaranteed to reproduce atomistic rdf's. This is because the calculated projections of the CG force field, $G_\zeta(x) \odot F$, weight the contributions of each force function, $\phi_\zeta(x)$, according to $G_{\zeta\zeta}$ (see eq 8). Equation 9 demonstrates that $G_{\zeta\zeta}$ is defined by an ensemble average according to the atomistic probability distribution or, equivalently, the many-body PMF, i.e., $G_{\zeta\zeta} \equiv G_{\zeta\zeta}[U^0]$. Thus, these force field projections are calculated by applying the MS-CG force field to configurations sampled by the atomistic model. However, when attempting to reproduce atomistic mean forces, it is not the many-body correlations in the atomistic model that are crucial but, rather, the many-body correlations generated by the CG model. Since simulations with the approximate CG force field will not generate a distribution of configurations identical to the many-body PMF, the MS-CG calculation may incorrectly estimate the contributions of correlated interactions to the CG mean forces. Consequently, MD simulations with the MS-CG force field may not perfectly reproduce the atomistic mean forces or, equivalently, rdf's.

These considerations suggest a self-consistent formulation of the MS-CG method for reproducing the atomistic mean forces. The CG potential, U^* , that reproduces a set of atomistic mean forces must satisfy eq 8 for b_ζ when the contribution of each force function is weighted according to $G_{\zeta\zeta}[U^*]$, that is, the corresponding ensemble average for U^* . Clearly, the normal MS-CG equations approximate this self-consistent equation by $G_{\zeta\zeta}[U^*] \approx G_{\zeta\zeta}[U^0]$. A more accurate estimate of $G_{\zeta\zeta}[U^*]$ should lead to a more accurate estimate of U^* and a more accurate reproduction of atomistic rdf's. This self-consistent

condition underlies the iterative-YBG framework developed by Cho and Chu.⁴⁹ In their approach, the normal MS-CG equations are iteratively solved, each time using a new estimate of $G_{\zeta\zeta}[U^*]$ that is determined by the previous approximate CG potential. In this context, it is particularly intriguing that the eigenvalues of $G_{\zeta\zeta}[U^0]$ are often bounded by 1, which may play an important role in the accuracy and robustness of the MS-CG method as well as the relative insensitivity of rdf's to many features of CG potentials.²¹ We also note that Weeks and co-workers have developed a considerably different self-consistent framework for determining short-ranged effective potentials that is also based upon the YBG equation.^{84,85}

Finally, we anticipate that the present development of the g-YBG framework may prove useful for quantifying and systematically improving the approximations underlying knowledge-based CG protein models.⁸⁶ Knowledge-based methods often determine potentials between pairs of amino acids based upon the frequency of observing the pair at a given distance in protein databank structures.^{87–89} The contributions of the surrounding environment to the corresponding mean force are typically estimated by a “reference state,” although the precise definition of this reference state remains somewhat unclear.^{90,91} Mullinax and Noid⁹² have demonstrated that, by directly addressing these contributions, an extended ensemble⁹³ version of the g-YBG theory^{25,26} can quantitatively recover the underlying potential from a model protein databank generated with a popular protein potential.⁹⁴ In this context, eq 12 provides a quantitative framework for identifying the contribution of this reference state to the mean force. Future work will investigate these contributions in the context of protein structures.

SUMMARY AND CONCLUSIONS

The present work reports a detailed analysis of many-body correlations in determining CG models that accurately reproduce atomistic distribution functions. In particular, these many-body correlations enter into the MS-CG and g-YBG approaches as a metric tensor that defines the angle between force field vectors on the basis of correlations between the corresponding interactions in the atomistic model. This metric tensor decomposes atomistic mean forces into direct and correlated contributions from the CG force field. Our calculations for liquid heptane demonstrate that this metric tensor reflects relatively generic features of molecular packing. Moreover, our calculations demonstrate that correlated interactions make significant contributions to these mean forces, although these mean forces appear relatively insensitive to many features of the CG force field.

Our results suggest several future directions. Since the present results correspond to a relatively simple system with weak alignment properties, it will be interesting to further consider this framework for more complex systems. In particular, we anticipate that the MS-CG metric tensor may prove useful for quantifying the interactions and correlations that are important for stabilizing complex molecular structures. In addition, the g-YBG framework may prove useful for characterizing and improving the approximations inherent to knowledge-based protein potentials. More generally, we anticipate that improved understanding of many-body correlations will contribute to developing CG potentials that more accurately model both structural and thermodynamic properties of complex condensed-phase systems.

■ ASSOCIATED CONTENT

§ Supporting Information

Detailed description of all methods, simulations, and calculations employed in this work as well as additional analysis of these calculations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: wnoid@chem.psu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The present work has been financially supported by an NSF CAREER award (Grant No. MCB 1053970) from the National Science Foundation and also by start-up funds from The Pennsylvania State University. J.F.R. was also funded in part by the Penn State Academic Computing Fellowship. Numerical calculations were performed using support and resources from Research Computing and Cyberinfrastructure, a unit of Information Technology Services at Penn State. Figures 2, 3, and 4 were made with VMD. VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

■ REFERENCES

- (1) *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; CRC Press: Boca Raton, FL USA, 2009.
- (2) Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869–1892.
- (3) Peter, C.; Kremer, K. *Faraday Discuss.* **2010**, *144*, 9–24.
- (4) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.
- (5) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464–4470.
- (6) Shinoda, W.; Devane, R.; Klein, M. L. *Mol. Simul.* **2007**, *33*, 27–36.
- (7) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.
- (8) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (9) Srinivas, G.; Nielsen, S.; Moore, P.; Klein, M. *J. Am. Chem. Soc.* **2006**, *128*, 848–853.
- (10) Rzepiela, A. J.; Sengupta, D.; Goga, N.; Marrink, S. J. *Faraday Discuss.* **2010**, *144*, 431–443.
- (11) Maerzke, K. A.; Siepmann, J. I. *J. Phys. Chem. B* **2011**, *115*, 3452–3465.
- (12) Rossi, G.; Monticelli, L.; Puisto, S. R.; Vattulainen, I.; Ala-Nissila, T. *Soft Matter* **2011**, *7*, 698–708.
- (13) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (14) DeVane, R.; Shinoda, W.; Moore, P. B.; Klein, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 2115–2124.
- (15) Liwo, A.; Oldziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849–873.
- (16) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (17) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.
- (18) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323–2347.
- (19) Lifson, S.; Oppenheim, I. *J. Chem. Phys.* **1960**, *33*, 109–115.
- (20) Hill, T. L. *Statistical Mechanics: Principles and Selected Applications*; Dover: New York, 1987.
- (21) Müller-Plathe, F. *ChemPhysChem* **2002**, *3*, 754–769.
- (22) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730–3737.
- (23) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.
- (24) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.
- (25) Mullinax, J. W.; Noid, W. G. *Phys. Rev. Lett.* **2009**, *103*, 198104.
- (26) Mullinax, J. W.; Noid, W. G. *J. Phys. Chem. C* **2010**, *114*, 5661–5674.
- (27) Mullinax, J. W.; Noid, W. G. *J. Chem. Phys.* **2010**, *133*, 124107.
- (28) Ellis, C. R.; Rudzinski, J. F.; Noid, W. G. *Macromol. Theory Simul.* **2011**, *20*, 478–495.
- (29) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Addison Wesley Publishing Company: Reading, MA, 1997.
- (30) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: San Diego, CA, USA, 1990.
- (31) Dias, C. L.; Ala-Nissila, T.; Grant, M.; Karttunen, M. *J. Chem. Phys.* **2009**, *131*, 054505.
- (32) Warren, P. *Phys. Rev. E* **2003**, *68*, 066702.
- (33) Baschnagel, J.; Binder, K.; Doruker, P.; Gusev, A. A.; Hahn, O.; Kremer, K.; Mattice, W. L.; Muller-Plathe, F.; Murat, M.; Paul, W.; Santos, S.; Suter, U. W.; Tries, V. *Adv. Polym. Sci.* **2000**, *152*, 41–156.
- (34) Louis, A. A.; Bolhuis, P. G.; Hansen, J. P.; Meijer, E. J. *Phys. Rev. Lett.* **2000**, *85*, 2522–2525.
- (35) Bolhuis, P.; Louis, A.; Hansen, J.; Meijer, E. J. *Chem. Phys.* **2001**, *114*, 4296–4311.
- (36) Schommers, W. *Phys. Rev. A* **1983**, *28*, 3599–3605.
- (37) Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624–1636.
- (38) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1997**, *55*, 5689–5696.
- (39) Lyubartsev, A.; Mirzoev, A.; Chen, L. J.; Laaksonen, A. *Faraday Disc.* **2010**, *144*, 43–56.
- (40) Savelyev, A.; Papoian, G. A. *J. Phys. Chem. B* **2009**, *113*, 7785–7793.
- (41) Savelyev, A.; Papoian, G. A. *Biophys. J.* **2009**, *96*, 4044–4052.
- (42) Savelyev, A.; Papoian, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20340–20345.
- (43) Murtola, T.; Karttunen, M.; Vattulainen, I. *J. Chem. Phys.* **2009**, *131*, 055101.
- (44) Shell, M. S. *J. Chem. Phys.* **2008**, *129*, 144108.
- (45) Chaimovich, A.; Shell, M. S. *Phys. Rev. E* **2010**, *81*, 060104.
- (46) Chaimovich, A.; Shell, M. S. *J. Chem. Phys.* **2011**, *134*, 094112.
- (47) Noid, W. G.; Liu, P.; Wang, Y. T.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.
- (48) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4116–4127.
- (49) Cho, H. M.; Chu, J. W. *J. Chem. Phys.* **2009**, *131*, 134107.
- (50) Krishna, V.; Larini, L. *J. Chem. Phys.* **2011**, *135*, 124103.
- (51) Rudzinski, J. F.; Noid, W. G. *J. Chem. Phys.* **2011**, *135*, 214101.
- (52) Ruhle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theory Comput.* **2009**, *5*, 3211–3223.
- (53) Das, A.; Andersen, H. C. *J. Chem. Phys.* **2009**, *131*, 034102.
- (54) Krishna, V.; Noid, W. G.; Voth, G. A. *J. Chem. Phys.* **2009**, *131*, 024103.
- (55) Lu, L. Y.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 954–965.
- (56) Das, A.; Andersen, H. C. *J. Chem. Phys.* **2010**, *132*, 164106.
- (57) Larini, L.; Lu, L. Y.; Voth, G. A. *J. Chem. Phys.* **2010**, *132*, 164107.
- (58) Izvekov, S.; Chung, P. W.; Rice, B. M. *J. Chem. Phys.* **2010**, *133*, 064109.
- (59) Lu, L.; Voth, G. A. *J. Chem. Phys.* **2011**, *134*, 224107.
- (60) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (61) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

- (62) Nose, S. *Mol. Phys.* **1984**, *52*, 255–268.
(63) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(64) Parrinello, M.; Rahman, A. *J. Chem. Phys.* **1982**, *76*, 2662–2666.
(65) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *99*, 8345–8348.
(66) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(67) Papaloannoy, D.; Zlakas, D.; Panayiotou, C. *J. Chem. Eng. Data* **1991**, *36*, 35–39.
(68) Berendsen, H.; Grigera, J.; Straatsma, T. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
(69) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
(70) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*; Cambridge University Press: New York, NY USA, 1992.
(71) Williams, T.; Kelley, C. *Gnuplot 4.4: an interactive plotting program*; <http://gnuplot.sourceforge.net/>, 2010.
(72) Anderson, E.; Bai, Z.; Bischof, C.; Blackford, S.; Demmel, J.; Dongarra, J.; Croz, J. D.; Greenbaum, A.; Hammarling, S.; McKenney, A. et al. *LAPACK Users' Guide*; SIAM: Philadelphia, PA USA, 1999.
(73) Harmandaris, V. A.; Adhikari, N. P.; van der Vegt, N. F. A.; Kremer, K. *Macromolecules* **2006**, *39*, 6708–6719.
(74) Soper, A. K. *Chem. Phys.* **1996**, *202*, 295–306.
(75) Izvekov, S. *J. Chem. Phys.* **2011**, *134*, 034104.
(76) Cho, H. M.; Gross, A. S.; Chu, J.-W. *J. Am. Chem. Soc.* **2011**, *133*, 14033–14041.
(77) Andersen, H. C.; Chandler, D.; Weeks, J. D. *Adv. Chem. Phys.* **1976**, *34*, 105.
(78) Chu, J.-W.; Izvekov, S.; Voth, G. A. *Mol. Simul.* **2006**, *32*, 211–218.
(79) Izvekov, S.; Violi, A.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 17019–17024.
(80) Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566–11575.
(81) Lu, L.; Voth, G. A. *J. Phys. Chem. B* **2009**, *113*, 1501–1510.
(82) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2009**, *113*, 4443–4455.
(83) Wang, Y. T.; Voth, G. A. *J. Am. Chem. Soc.* **2005**, *127*, 12192–12193.
(84) Chen, Y. G.; Kaur, C.; Weeks, J. D. *J. Phys. Chem. B* **2004**, *108*, 19874–19884.
(85) Rodgers, J. M.; Kaur, C.; Chen, Y. G.; Weeks, J. D. *Phys. Rev. Lett.* **2006**, *97*, 097801.
(86) Skolnick, J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 166–171.
(87) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 945–950.
(88) Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, *18*, 534–552.
(89) Sippl, M. *J. Mol. Biol.* **1990**, *213*, 859–883.
(90) Thomas, P. D.; Dill, K. A. *J. Mol. Biol.* **1996**, *257*, 457–469.
(91) Betancourt, M. R. *Proteins* **2009**, *76*, 72–85.
(92) Mullinax, J. W.; Noid, W. G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19867–19872.
(93) Mullinax, J. W.; Noid, W. G. *J. Chem. Phys.* **2009**, *131*, 104110.
(94) Honeycutt, J. D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3526–3529.