Published in final edited form as:
J Phys Chem B. 2007 April 26; 111(16): 4116–4127. doi:10.1021/jp068549t.

Multiscale Coarse-graining and Structural Correlations: Connections to Liquid State Theory

W. G. Noid, **Jhih-Wei Chu**, **Gary S. Ayton**, and **Gregory A. Voth***
Center for Biophysical Modeling and Simulation and Department of Chemistry, University of Utah, 315 S. 1400 E. Rm 2020, Salt Lake City, Utah 84112-0850

Abstract

A statistical mechanical framework elucidates the significance of structural correlations between coarse-grained (CG) sites in the multiscale coarse-graining (MS-CG) method [S. Izvekov and G.A. Voth. J. Phys. Chem. B **109** 2469 (2005), J. Chem. Phys. **123** 134105 (2005)]. If no approximations are made, the MS-CG method yields a many-body multi-dimensional potential of mean force describing the interactions between CG sites. However, numerical applications of the MS-CG method typically employ a set of pair potentials to describe non-bonded interactions. The analogy between coarse-graining and the inverse problem of liquid state theory clarifies the general significance of three-particle correlations for the development of such CG pair potentials. It is demonstrated that the MS-CG methodology incorporates critical three-body correlation effects and that, for isotropic homogeneous systems evolving under a central pair potential, the MS-CG equations are a discretized representation of the well-known Yvon-Born-Green equation. Numerical calculations validate the theory and illustrate the role of these structural correlations in the MS-CG method.

1. Introduction

Although molecular dynamics (MD) simulations provide a powerful tool for investigating complex biomolecular systems,[1] their substantial computational cost limits conventional atomistic MD simulations to investigations on time-scales that are less than microseconds and length-scales that are significantly less than micrometers.[2,3] Such atomistic MD simulations are often inadequate to model biological processes such as protein folding[4] or signal transduction,[5] which may occur on significantly larger time- and length- scales. Consequently, there has been growing interest in developing computationally-inexpensive "coarse-grained" (CG) models,[6–18] which can then be simulated over significantly longer time- and length- scales.

In order to reproduce the thermodynamic and structural properties of the original atomistic system, the low-resolution CG structures must be sampled according to the probability distribution function for the fully atomistic representation of the same structures. A formal prescription for designing CG models thus involves the integration over atomistic degrees of freedom to define a reduced description of the system.[2,7] The interactions between CG sites must therefore include not only energetic, but also entropic contributions that result from averaging over eliminated degrees of freedom. In principle, the resulting CG model may require a many-body interaction potential that depends upon the thermodynamic state point.[13] In practice, though, CG force fields typically model non-bonded interactions with central pair potentials that depend only upon the distance between CG sites.[2] This effective pair potential represents an approximate decomposition of the many-body interaction obtained from a formal

*Electronic mail: voth@chemistry.chem.utah.edu.

integration over uninteresting degrees of freedom. The procedure for determining this pair potential must appropriately incorporate the effects of many-body correlations in order to reproduce the structure of the original system.

The theory of coarse-grained modeling is similar to the 'inverse problem' of liquid state theory.[19–21] Both theories attempt to determine an interaction potential reproducing an observed structure. However, in coarse-grained modeling the target structure is a low-resolution representation of the original structure and the deduced interaction is between the CG sites defining the reduced structure.[2] The theory of the Yvon-Born-Green (YBG) equation provides a direct solution to this inverse problem,[20] under the assumption that such an interaction potential exists. The YBG equation provides an exact relation between a given two-body interaction potential and the n- and (n+1)- particle distribution functions obtained from equilibrium simulations employing the potential. Therefore, a CG pair potential may be determined by inverting the YBG equation for the observed two- and three-particle CG distribution functions. This relationship suggests a role for higher order correlations in deducing a pair potential that will reproduce the observed CG structure.

The multiscale coarse-graining (MS-CG) procedure[16,17] determines CG force fields from atomistic MD simulations by employing a statistical implementation of the Force-Matching (FM) method.[22,23] This MS-CG method has been successful in developing CG models for many complex systems such as ionic liquids,[24] mixed lipid bilayers,[25] small peptides,[26] nano-particles, [27] and even mixed resolution models of trans-membrane proteins in lipid bilayers.[28] In the following analysis it is demonstrated that if no approximations in the functional form of the CG force field are made, the MS-CG method determines a many-body multidimensional potential of mean force describing the CG representation of the system. Consequently, simulations employing this many-body MS-CG potential would generate CG structures according to the underlying atomistic distribution function. However, prior numerical applications of the MS-CG method have approximated this many-body interaction potential with a set of bonded and non-bonded pair potentials between CG sites.[16,17,24–28] The present work demonstrates that the MS-CG equations for this CG pair potential reflect both two- and three-particle correlations between CG sites within a system. Moreover, for homogeneous, isotropic systems the MS-CG equations are equivalent to generalized Yvon-Born-Green (YBG) equations[20] for CG sites interacting according to a central pair force field. For such systems the MS-CG methodology explicitly considers the two- and three- particle correlations between CG sites within an atomistic MD simulation, assumes that these distribution functions were generated by a pair-wise decomposable central force field, and then inverts the resulting YBG equation to determine this force field. The YBG theory thus provides a fundamental link between the system structure and the CG force field.

In Section 2.1 the relationship between the MS-CG interaction potential and a multidimensional potential of mean force is derived. The "normal" MS-CG equations are next derived in Section 2.2 by approximating the many-body interaction potential with a central pair-wise decomposable force field. The derivation emphasizes the relationship between the MS-CG equations and the correlations observed between CG sites in atomistic MD simulations. The YBG equation for a CG system is then presented in Section 2.3 and it is shown that for a homogeneous, isotropic system this equation reduces to the MS-CG equations. Numerical illustrations of this analysis are presented in Section 3. In Section 4 the significance of liquid state theory for developing coarse-grained models is considered, especially with regards to the MS-CG and reverse Monte Carlo methods.[2,21,29] These results are reviewed in the Conclusion section, Section 5. Proofs of certain necessary results and generalizations of the theory for a multi-component system are provided in the Appendices. A more detailed discussion of the theory is attached as a Supporting Information section.

2. Theory

2.1. Many-body CG potential of mean force

The multi-scale coarse-graining (MS-CG) method of Izvekov and Voth extends the Force-Matching (FM) method[22] to determine coarse-grained (CG) force fields from atomistic molecular dynamics simulations.[17] The CG force field is obtained by minimizing a residual describing the difference between the instantaneous forces defined by the CG force field and the original atomistic force field. This difference is statistically averaged over all CG sites and all configurations sampled from the atomistic MD simulations. As shown below, these configurations must be sampled according to the distribution function for the atomistic Hamiltonian in order to perform the correct averaging. The FM residual for a system described by N_{CG} identical CG sites may be expressed:

$$\chi^2=\frac{1}{3N_I N_{CG}}\sum_I^{N_I}\sum_i^{N_{CG}}\left|\vec{F}_i^{I,AA}-\vec{F}_i^{I,CG}\right|^2. \tag{1}$$

In eq (1) and in the following, Latin subscripts (*i*) indicate particular CG sites and the superscript (*I*) labels configurations sampled during the atomistic MD simulation. Thus, the MS-CG residual compares the total force on a given CG site defined by the atomistic force field for the sampled atomistic configuration, $\vec{F}_i^{I,AA}$, with the force on the same site defined by the CG force field for the CG representation of the same configuration, $\vec{F}_i^{I,CG}$.

In the limit of adequate sampling, the FM residual may be considered an average over the region of configuration space, *D*, sampled by an atomistic trajectory and weighted according to the atomistic distribution function, $\rho_{AA}(\vec{r}^{N_{AA}})$, defined for the trajectory. The residual may then be expressed as the configurational integral:

$$\chi^2=\frac{1}{3N_{CG}}\int_D d\vec{r}^{N_{AA}}\rho_{AA}(\vec{r}^{N_{AA}})\sum_i\left|\vec{F}_i^{AA}-\vec{F}_i^{CG}\right|^2. \tag{2}$$

Now, assume that there exists a canonical transformation that partitions the atomistic phase space into a set of coordinates describing the CG sites and a set of residual degrees of freedom: $\vec{r}^{N_{AA}}=(\vec{r}^{N_{CG}},\vec{r}^{N_R})$. Such a transformation certainly exists, for example, when the CG sites are defined as the centers of mass for groups of atoms. Upon employing this transformation in eq (2), the residual may be considered a functional of the CG force field:

$$\chi^2\left[\vec{F}^{CG}(\vec{r}^{N_{CG}})\right]=\frac{1}{3N_{CG}}\int_{D_{CG}}d\vec{r}^{N_{CG}}\int_{D_R}d\vec{r}^{N_R}\rho_{AA}\left(\vec{r}^{N_{AA}}\right)\sum_i\left|\vec{F}_i^{AA}(\vec{r}^{N_{AA}})-\vec{F}_i^{CG}(\vec{r}^{N_{CG}})\right|^2. \tag{3}$$

The CG force field is defined by minimizing the functional[20,30] according to $\delta\chi^2\left[\vec{F}^{CG}(\vec{r}^{N_{CG}})\right]/\delta\vec{F}_i^{CG}(\vec{r}^{N_{CG}})=0$, yielding the result:

$$\vec{F}_i^{CG}(\vec{r}^{N_{CG}})=e^{+\beta V^{CG}(\vec{r}^{N_{CG}})}\int_{D_R}d\vec{r}^{N_R}e^{-\beta V^{AA}(\vec{r}^{N_{CG}},\vec{r}^{N_R})}\vec{F}_i^{AA}\left(\vec{r}^{N_{CG}},\vec{r}^{N_R}\right). \tag{4}$$

Here the canonical ensemble is explicitly considered: $\rho_{AA}(\vec{r}^{N_{AA}})=e^{-\beta V^{AA}(\vec{r}^{N_{AA}})}/Z(N,V_0,T)$ where $Z(N,V_0,T)$ is the canonical partition function. The force on CG site i according to the atomistic potential is defined $\vec{F}_i^{AA}(\vec{r}^{N_{AA}})=-\partial V^{AA}(\vec{r}^{N_{CG}},\vec{r}^{N_R})/\partial\vec{r}_i$, where the partial derivative is performed with respect to the CG coordinate $\vec{r}_i$ while holding all remaining atomistic and CG coordinates fixed. In general the force $\vec{F}_i^{AA}$ depends upon all $N_{AA} = N_{CG} + N_R$ degrees of freedom. The CG force field defined by eq (4) depends upon all N_{CG} degrees of freedom and may be considered an average of the atomistic force acting on a CG site with the average performed over all sampled atomistic configurations consistent with the fixed CG configuration and weighted according to the atomistic distribution function. The normalization of eq (4) defines a multi-dimensional potential of mean force (pmf) describing the CG sites that may be expressed:

$$e^{-\beta V^{CG}(\vec{r}^{N_{CG}})}=\int_{D_R} d\vec{r}^{N_R} e^{-\beta V^{AA}(\vec{r}^{N_{CG}},\vec{r}^{N_R})}. \tag{5}$$

The CG force field defined in eq (4) is the appropriate gradient of this CG potential:

$$\vec{F}_i^{CG}(\vec{r}^{N_{CG}})=-\left(\frac{\partial}{\partial\vec{r}_i}V^{CG}(\vec{r}^{N_{CG}})\right)_{\vec{r}^{(N_{CG}-1)}}. \tag{6}$$

2.2. Force-Matching to a Central Pair Potential

Previous numerical applications of the MS-CG method[16,17,24–27] have approximated the many-body CG potential of mean force (pmf) derived above in eq (5) with a central pair potential. These applications have determined the optimal CG force field by minimizing the residual in eq (1) under the assumptions that the MS-CG force field is pair-wise additive, such that

$$\vec{F}_i^{I,CG}=\sum_{j\neq i}\vec{F}_{i,j}^{CG}\left(\vec{r}_i^I,\vec{r}_j^I\right) \tag{7}$$

and, furthermore, that this force field is central,

$$\vec{F}_{i,j}^{CG}\left(\vec{r}_i^I,\vec{r}_j^I\right)=\vec{u}_{i,j}^I f\left(r_{i,j}^I\right). \tag{8}$$

In these definitions $\vec{r}_i$ represents the Cartesian coordinates of CG site i; $\vec{r}_{i,j}=\vec{r}_i-\vec{r}_j$ represents the vector from the j to the i CG site; $\vec{r}_{i,j}=\left|\vec{r}_i-\vec{r}_j\right|$ represents the magnitude of this vector; $\vec{u}_{i,j}=\vec{r}_{i,j}/\vec{r}_{i,j}$ represents the associated unit vector; and $f(r)$ represents the function defining the magnitude of the interaction between CG sites and depends only upon the inter-site distance. The force field, $f(r)$, minimizing the residual under these constraints is uniquely determined and may be conveniently described by a discrete delta function basis for which $\delta_D(r)=1$ when $-\Delta r/2\leq r<\Delta r/2$ and is 0 otherwise. In this basis the force field may be represented as:

$$f(r)=\sum_{d}^{N_d} f_d \delta_D(r - r_d). \tag{9}$$

This definition corresponds to tabulating the force field at a discrete set of points, r_d, about which ($r_d - \Delta r / 2 < r < r_d + \Delta r / 2$) the force field is assumed to be constant. Previous applications of the MS-CG method[16,17,24–27] have typically employed a spline basis for representing the FM force field. However, the basis used in eq (9) is particularly amenable for the following analysis and, in the limit that $\Delta r \to 0$, this representation transparently recovers a continuous representation of the force field.

By minimizing the residual (1) with respect to the elements of the force table, $\partial \chi^2 / \partial f_d = 0$, and employing the definitions in eq (7)–eq (9), a system of linear algebraic equations is obtained:

$$\sum_{d'} f_{d'} G_{dd'} = b_d, \tag{10}$$

in which

$$b_d = \left\langle \sum_i \sum_{j \neq i} (\vec{F}_i^{I,AA} \cdot \vec{u}_{i,j}^{I}) \delta_D(r_{i,j}^I - r_d) \right\rangle_I \tag{11}$$

and

$$G_{dd'} = \left\langle \sum_i \sum_{j \neq i} \sum_{k \neq i} (\vec{u}_{i,j}^{I} \cdot \vec{u}_{i,k}^{I}) \delta_D(r_{i,j}^I - r_d) \delta_D(r_{i,k}^I - r_{d'}) \right\rangle_I . \tag{12}$$

The angular brackets denote an average over configurations sampled by the atomistic MD simulation. In the following analysis it will be assumed that this average over configurations is equivalent to the appropriate ensemble average:

$$\langle A(r) \rangle_I = \frac{1}{N_I} \sum_I A^I(r) = \int d\vec{x}^N \rho\left(\vec{x}^N\right) A(r; \vec{x}^N) = \langle A(r) \rangle. \tag{13}$$

The linear system in eq (10) is referred to as the “normal” MS-CG equations because $G_{dd'}$ is a symmetric, i.e. normal, matrix. Previous numerical implementations of the MS-CG procedure[16,17,24–28] have employed an additional block-averaging approximation to solve an equivalent system of over-determined equations for the force field, f_d. It is the purpose of this work to further elucidate the physical significance of the “normal” MS-CG equations and, in particular, to relate them to well-known theories for the liquid state.[20]

In eq (10) all information regarding the atomistic forces has been packaged into the term b_d. According to eq (11), this information enters as the average projection of the instantaneous total force on each CG site, $\vec{F}_i^{I,AA}$, onto the sum of unit vectors from all other CG sites, $\vec{u}_{i,j}^{I}$, that are a distance r_d from i in the given configuration, I. Thus, the quantity b_d describes the average correlation between the instantaneous net force on each CG site and the local spatial

distribution of CG sites a distance r_d away. If this distribution is always instantaneously spherically symmetric, then $b_d = 0$. Similarly, $G_{dd'}$ describes the average instantaneous fluctuations in the local density of CG sites at distances r_d and $r_{d'}$, respectively, from each CG site.

Since the $j = k$ term has not been excluded from the triple sum in eq (12), the quantity $G_{dd'}$ includes both two- and three-particle correlations, which may be explicitly separated and analyzed:

$$G_{dd'}=\delta_{dd'}G_d^{(2)}+G_{dd'}^{(3)} \tag{14}$$

$$G_d^{(2)}=\langle\sum_i\sum_{j\neq i}\delta_D(r_{i,j}^I-r_d)\rangle_I \tag{15}$$

$$G_{dd'}^{(3)}=\langle\sum_i\sum_{j\neq i,k\neq i,j}\sum(\vec{u}_{i,j}^I\cdot\vec{u}_{i,k}^I)\delta_D(r_{i,j}^I-r_d)\delta_D(r_{i,k}^I-r_{d'})\rangle_I . \tag{16}$$

The quantity $G_d^{(2)}$ counts the number of distinct CG sites separated by a distance r_d and is closely related to the radial distribution function for the CG sites. The quantity $G_{dd'}^{(3)}$ is a direct measure of three-particle correlations between CG sites in atomistic MD simulations. The factor $\vec{u}_{i,j}^I\cdot\vec{u}_{i,k}^I=\cos\theta_{i;j,k}$ is the cosine of the angle defined between the three CG sites with the site i at the vertex of the angle as illustrated in Figure 1. Thus, $G_{dd'}^{(3)}$ describes the constrained average value of this quantity for two distinct CG sites, j and k, that are distances r_d and $r_{d'}$, respectively, from each CG site, i. Because the three CG sites are distinct, excluded volume effects prevent sites j and k from overlapping. Consequently, for $r_d \approx r_{d'}$, there exists a cone of small angles defined by $\theta_{i;j,k} \leq \theta_{i;j,k}^* \approx 0$, (such that $\vec{u}_{i,j}^I\cdot\vec{u}_{i,k}^I \approx 1$,) that are never sampled during the MD simulation. Corresponding configurations for which $\theta_{i;j,k} \approx \pm\pi$ (and $\vec{u}_{i,j}^I\cdot\vec{u}_{i,k}^I \approx -1$) are not disfavored and, as a result, the constrained average is negative for distances $r_d \approx r_{d'}$ at which small angle configurations are disfavored. This situation is illustrated in Figure 1b. The form of $G_{dd'}^{(3)}$ may at first seem somewhat artificial. The dot product factor arises as a consequence of the assumption in eq (8) that the MS-CG force field is directed along the inter-site vector. Moreover, it will be shown that this same factor arises quite naturally in liquid state theories for a central pair potential.

Employing the definitions in eq (14)–eq (16), eq (10) may be re-expressed as:

$$b_d - f_d G_d^{(2)}=\sum_{d'} f_{d'}G_{dd'}^{(3)}. \tag{17}$$

As mentioned above, all information regarding atomistic forces is contained within b_d. The left hand side of eq (17) depends only upon two-particle information, while the right hand expression reflects three-particle correlations through the term $G_{dd'}^{(3)}$, which couples the

equations for different force table elements. The first term in the left hand expression (i.e., b_d) describes the average correlation of the total instantaneous force on each CG particle with the spatial distribution of CG particles a distance r_d away. The second term describes the average net force on each CG particle from CG particles a distance r_d away in terms of the MS-CG force field and the average two-particle distribution according to $f_d G_d^{(2)}$. The difference between this average projection of the total force and the average force arising directly from a CG particle at the fixed distance arises from interactions with a third particle. The right hand side of eq (17) then describes how forces from a third particle are correlated with the two-particle structure described in $G_d^{(2)}$. The quantity $f_{d'} G_{dd'}^{(3)}$ describes the average MS-CG force from a distinct third particle, *k*, a distance $r_{d'}$ away from the *i* CG site, given that the *j* CG site is a distance r_d away. The dot product in the definition of $G_{dd'}^{(3)}$ arises because it has been assumed that the interaction between each pair is along the vector connecting them.

Although the numerical MS-CG procedure explicitly employs information regarding the atomistic forces, it is proven in Appendix A that the normal MS-CG equations presented in eq (10) may be recast in a form that is independent of atomistic force information. Under the assumption that the forces on the CG sites measured in MD simulations may be expressed as the gradient of a many-body CG potential energy function generating the observed CG structure, it follows that:

$$\frac{dg(r)}{dr} = \frac{1}{k_B T}\left(\frac{1}{4\pi\rho^2 V_0}\right)\frac{1}{r^2}\left\langle \sum_i \sum_{j\neq i} (\vec{F}_i^{I,AA} \cdot \vec{u}_{i,j}^{I}) \delta_D(r_{i,j}^I - r)\right\rangle_I , \tag{18}$$

where

$$g(r) = \left(\frac{1}{4\pi\rho^2 V_0}\right)\frac{1}{r^2}\left\langle \sum_i \sum_{j\neq i} \delta_D(r_{i,j}^I - r)\right\rangle_I . \tag{19}$$

In eq (18) and eq (19) it has been assumed that the average over MD configurations is equivalent to the ensemble average, $\rho = N/V_0$, and V_0 is the total system volume. Equation (19) defines the radial distribution function for CG sites in an atomistic MD simulation.[20] As proven in Appendix A, eq (18) follows immediately from the definition in eq (19).

Multiplying both sides of eq (17) by the factor $1/(4\pi\rho^2 V_0 \cdot r_d^2)$ and employing the relations in eq (18) and eq (19), the MS-CG equations for the interaction, f_d, may be expressed as:

$$k_B T \frac{dg}{dr}\Big|_{r_d} - f_d\, g\big|_{r_d} = \sum_{d'}\left[f_{d'}\left(\frac{1}{4\pi\rho^2 V_0}\right)\frac{1}{r_d^2} G_{dd'}^{(3)}\right]. \tag{20}$$

Equation (20) contains the same information as eq (17) but has eliminated the explicit dependence upon atomistic forces for an equivalent quantity in terms of the radial distribution function. This equation emphasizes the discrete nature of the MS-CG equations and suggests a continuous representation. The discrete delta function representation defined in eq (9) is particularly convenient for this purpose as, in the limit that $\Delta r \to 0$, eq (20) becomes a one-dimensional linear integral equation

$$k_B T \frac{dg(r)}{dr} - f(r)g(r) = \int dr' f(r') M^{FM}(r,r'), \tag{21}$$

in terms of a FM kernel which describes the effects of three-particle correlations according to

$$\begin{aligned} M^{FM}(r,r') &= \left(\frac{1}{4\pi\rho^2 V_0}\right) \frac{1}{r^2} G^{(3)}(r,r') \\ &= \left(\frac{1}{4\pi\rho^2 V_0}\right) \frac{1}{r^2} \left\langle \sum_i \sum_{j\neq i, k\neq i,j} \sum (\vec{u}^I_{i,j} \cdot \vec{u}^I_{i,k}) \delta(r^I_{i,j} - r) \delta(r^I_{i,k} - r') \right\rangle_I \end{aligned} \tag{22}$$

Equation (21) re-emphasizes the significance of three-particle correlations in the MS-CG procedure. The left hand side of this equation depends only upon two-particle information and, in the case that the right hand side vanishes, the MS-CG interaction becomes $f(r) = k_B T \, d(\ln g(r))/dr = -w'(r)$, where $w(r) = -k_B T \ln g(r)$, i.e., the conventional 2-body pmf. Thus the MS-CG pair potential differs significantly from the 2-body pmf because it explicitly considers the effects of three-particle correlations between CG sites in determining a CG force field.

The normal MS-CG equations for a multi-component system are derived in Appendix C. The resulting system of equations for the MS-CG interaction between CG sites of types α and β may be expressed

$$\left(k_B T \frac{d}{dr} - f^{\alpha\beta}(r)\right) g_{\alpha\beta}(r) = \sum_\gamma \int dr' \tfrac{1}{2} \left[M^{FM}_{\alpha;\beta\gamma}(r,r') f^{\alpha\gamma}(r') + M^{FM}_{\beta;\alpha\gamma}(r,r') f^{\beta\gamma}(r') \right]. \tag{23}$$

This equation generalizes eq (21) to consider multiple types of CG particles (labeled with Greek indices), the relevant interactions between them, and also the effect of three-particle correlations centered about the second particle, which is a distinct case for $\alpha \neq \beta$. A more detailed presentation of the derivation is provided in the Supporting Information.

To summarize the preceding analysis, the MS-CG method has been applied to derive a linear integral equation for the MS-CG force field. By assuming that the CG force field is both pair-wise additive and central, according to eq (7) and eq (8), and employing the discrete delta function basis of eq (9), the linear least squares problem defined by the residual in eq (1) has been transformed into a system of linear algebraic equations (10) describing the relationship between the CG force field and the distribution of CG sites. By separating two- and three-particle contributions according to eq (14) and employing the identity in eq (18) to eliminate atomistic force information, the normal MS-CG equations have been recast in a form that depends only upon structural information according to eq (20). Finally, using the discrete basis to pass into the continuum limit, this system of equations has been expressed as the integral equation in eq (21), which generalizes to eq (23) for systems with multiple types of CG sites.

2.3. Yvon-Born-Green Equation

As demonstrated in subsection 2.2, the normal MS-CG equations are a discrete representation of a one-dimensional linear integral equation. For a homogeneous isotropic system, this integral equation is equivalent to the Yvon-Born-Green equation[20] describing CG distribution functions resulting from a central pair force field. This result is briefly derived for a one-component system in the present subsection. The general result for multi-component systems is derived in Appendix D. A more detailed presentation is provided as Supporting Information.

Consider a system described by N_{CG} identical CG sites. The Hamiltonian for this system evolving under an N_{CG} -particle potential function, V^{NCG}, is defined:

$$H(\vec{p}^{N_{CG}}, \vec{r}^{N_{CG}}) = \sum_{i}^{N_{CG}} \frac{1}{2m} \left| \vec{p}_i \right|^2 + V^{N_{CG}}(\vec{r}^{N_{CG}}). \tag{24}$$

Assuming that the total force on each CG site, $\vec{F}_i = -\partial V^{N_{CG}} / \partial \vec{r}_i$, arises from a sum of pair interactions,

$$\vec{F}_i = \sum_j \vec{F}_{i,j}(\vec{r}_i, \vec{r}_j), \tag{25}$$

then for a homogeneous system with $\rho(\vec{r}_i) = \rho = \text{const}$, it follows that the distribution of three arbitrary, but distinct CG particles, *i*, *j*, and *k* may be described according to

$$\left(k_B T \frac{\partial}{\partial \vec{r}_i} - \vec{F}_{i,j}(\vec{r}_i, \vec{r}_j) \right) g^{(2)}(\vec{r}_i, \vec{r}_j) = \rho \int d\vec{r}_k \, \vec{F}_{i,k} g^{(3)}(\vec{r}_i, \vec{r}_j, \vec{r}_k). \tag{26}$$

Equation (26) is the well-known Yvon-Born-Green (YBG) equation[20] relating the equilibrium two- and three- CG particle distribution functions to the pair-wise decomposable force field, $\vec{F}_{i,j}$. If the system is also isotropic, the two-particle correlation function is equal to its translational and orientational average:

$$g^{(2)}\left(\vec{r}_i, \vec{r}_j\right) = \int \frac{1}{V_0} d\vec{r}_i \int \frac{1}{4\pi} d\Omega_{j,i} g^{(2)}\left(\vec{r}_i, \vec{r}_j\right) = \hat{\mathrm{P}}_{\mathrm{RT}} g^{(2)}\left(\vec{r}_i, \vec{r}_j\right) = g\left(r_{i,j}\right). \tag{27}$$

Equation (27) defines an operator $\hat{\mathrm{P}}_{\mathrm{RT}} = \int \frac{1}{V_0} d\vec{r}_i \int \frac{1}{4\pi} d\Omega_{j,i}$ which averages over both translation and rotation of the system, where $\Omega_{j,i}$ defines the orientation of the vector $\vec{r}_{j,i}$ from the *i* to the *j* CG site. For an isotropic homogeneous system the gradient term in eqs (26) may then be simplified as $\partial g^{(2)}\left(\vec{r}_i, \vec{r}_j\right) / \partial \vec{r}_i = \vec{u}_{i,j} \, dg\left(r_{i,j}\right) / dr_{i,j}$. Under the additional assumption that the pair interaction between CG sites is central such that,

$$\vec{F}_{i,j}\left(\vec{r}_i, \vec{r}_j\right) = \vec{u}_{i,j} f\left(r_{i,j}\right), \tag{28}$$

eq (26) may be re-expressed to read

$$\vec{u}_{i,j}\left(k_B T \frac{d}{dr_{i,j}} - f\left(r_{i,j}\right) \right) g\left(r_{i,j}\right) = \rho \int d\vec{r}_k \, \vec{u}_{i,k} f\left(r_{i,k}\right) g^{(3)}\left(\vec{r}_i, \vec{r}_j, \vec{r}_k\right). \tag{29}$$

Projecting this equation onto the vector $\vec{u}_{i,j}$ and shifting the integration variable, one obtains the following result:

$$\left(k_B T \frac{d}{dr_{i,j}} - f\left(r_{i,j}\right)\right) g\left(r_{i,j}\right) = \rho \int d\vec{r}_{k,i} f\left(r_{i,k}\right) \left(\vec{u}_{i,k} \cdot \vec{u}_{i,j}\right) g^{(3)}\left(\vec{r}_i, \vec{r}_j, \vec{r}_k\right). \tag{30}$$

Thus the dot product factor arises naturally in an integral equation theory, just as it did in the MS-CG equations. For a system described by a central pair potential without an external field, the average effect of a third particle on two-particle correlations must lie along the two-particle vector.

Because the left hand side of eq (30) depends only on the distance, $r_{i,j}$, upon application of the operator $\hat{\mathrm{P}}_{\mathrm{RT}}$ eq (30) may be expressed as a one-dimensional integral equation:

$$\left(k_B T \frac{d}{dr_{i,j}} - f\left(r_{i,j}\right)\right) g\left(r_{i,j}\right) = \int dr_{i,k} f\left(r_{i,k}\right) M^{YBG}\left(r_{i,j}, r_{i,k}\right). \tag{31}$$

Equation (31) defines a YBG kernel describing the effects of a third particle on two particle correlations,

$$M^{YBG}\left(r_{i,j}, r_{i,k}\right) = \rho\, r_{i,k}^2 \int d\Omega_{k,i} \hat{\mathrm{P}}_{\mathrm{RT}} \left(\vec{u}_{i,k} \cdot \vec{u}_{i,j}\right) g^{(3)}\left(\vec{r}_i, \vec{r}_j, \vec{r}_k\right). \tag{32}$$

Moreover, it is shown in Appendix B that

$$M^{YBG}\left(r_{i,j}, r_{i,k}\right) = M^{FM}\left(r_{i,j}, r_{i,k}\right) = M\left(r_{i,j}, r_{i,k}\right). \tag{33}$$

Therefore, for a homogeneous isotropic system evolving under a central pair potential, the YBG eq (26) may be reduced exactly to the FM equation (21):

$$\left(k_B T \frac{d}{dr} - f(r)\right) g(r) = \int dr'\, f(r') M(r, r'). \tag{34}$$

Appendix D generalizes the YBG theory for a homogeneous isotropic multi-component CG system. If the system is governed by a set of central pair potentials, the generalized YBG equation reduces to a form that is equivalent to the multi-component MS-CG equations (23):

$$\left(k_B T \frac{d}{dr} - f^{\alpha\beta}(r)\right) g_{\alpha\beta}(r) = \sum_{\gamma} \int dr' \tfrac{1}{2} \left[f^{\alpha\gamma}(r') M_{\alpha;\beta\gamma}(r, r') + f^{\beta\gamma}(r') M_{\beta;\alpha\gamma}(r, r')\right]. \tag{35}$$

A more detailed derivation is provided in the Supporting Information.

The analysis of section 2 elucidates the general significance of three-particle correlations for deducing CG pair potentials in general and the MS-CG pair force field in particular. Furthermore, the equivalence of eq (23) and eq (35) indicates that the MS-CG method incorporates higher-order correlations in a mechanism that is consistent with the well-known statistical mechanics of the liquid state.[20] However, while the analysis of subsection 2.1 and subsection 2.2 and, in particular, the “normal” MS-CG equations presented in eq (17), are generally valid and may be applied to determine a CG force field for any system, the preceding

analysis of subsection 2.3, which relates the MS-CG equations to the YBG theory, is strictly valid only for isotropic, homogeneous systems. Moreover, there exists no general proof that the two- and three- CG particle distribution functions measured from atomistic MD simulations may be related to a simple pair potential through a YBG-type equation.[19,31] Rather, as stressed in subsection 2.1, the residual defined by eq (1) is minimized by a many-body CG interaction potential. The MS-CG pair potential investigated in subsection 2.2 is only an approximate decomposition of this interaction potential, albeit one that has physical significance and which incorporates critical three-body correlations as demonstrated in subsection 2.2 and subsection 2.3.

3. Results

The analysis of the previous section demonstrates that the MS-CG equations reflect two- and three-particle correlations between CG sites observed in atomistic MD simulations. The following figures illustrate the effect of these structural correlations on the normal MS-CG equations for coarse-graining a system of 216 Lennard-Jones (LJ) spheres and a system of 216 simple point charge (SPC) water molecules.[32] The CG mapping for the LJ system is an identity operation and the resulting MS-CG equations reflect the atomistic structure. The SPC water system is coarse-grained onto two different one-site models: the COM (COG) model maps each SPC molecule onto a single site located at its center-of-mass (geometry). To facilitate comparison between these systems, the LJ sphere system was modeled using the LJ parameters for the oxygen atom in the SPC model. In the following figures, the solid lines correspond to the LJ system, the dashed lines correspond to the SPC system analyzed in terms of the COG, and the dotted lines correspond to the SPC system analyzed in terms of the COM. All MD simulations were performed in the constant NVT ensemble using the GROMACS 3.3.0 software package,[33] with V_0=(1.8602 nm)^3 and T = 298K maintained with the Nose-Hoover thermostat.[34,35] All quantities computed from MD simulation were tabulated on a grid with Δr = 0.001nm according to the discrete delta function basis defined in eq (9). Because the grid is so fine, the systems were simulated for 30ns in order to adequately converge all relevant quantities and accurately evaluate necessary numerical derivatives.

Analyzing the MS-CG procedure for the LJ fluid is instructive because the “coarse-graining” for this system is an identity mapping. Clearly there exists a central pair potential that will exactly reproduce the system structure (i.e., the LJ pair potential,) and it is illuminating to investigate the mechanism by which the MS-CG procedure recovers this pair potential from the relevant distribution functions. In contrast, there does not necessarily exist a pair potential for a one-site CG model of water that will reproduce both the two- and three- particle distribution functions measured in atomistic MD simulations.[19,31] The MS-CG procedure determines a central pair potential that would reproduce this structure under the assumption that the CG distribution functions were related by the YBG equation to the assumed central pair potential.

Equation (20) expresses the MS-CG equations in terms of the radial distribution function (rdf), $g(r)$. The rdf’s measured for CG sites in the atomistic simulations are presented in Figure 2. The maxima and minima of the LJ rdf are roughly evenly spaced, corresponding to simple packing. In contrast, the first peak of both SPC rdf’s is much more narrow and the first minima is at shorter range, reflecting the short-range packing effects in water due to the presence of hydrogen atoms. The SPC-COM rdf has a larger and more pronounced hard core region because the CG site roughly corresponds to the oxygen atom of each molecule and the hydrogen atoms that have been integrated out in the CG representation generate a larger excluded volume between the CG sites. In contrast, the SPC-COG model has a less well-defined hard-core region and presents a smaller excluded volume effect. The LJ rdf indicates much longer and more pronounced correlations than either representation of SPC water.

As indicated in section 2.2, if three-particle correlations were not considered in the MS-CG procedure, the resulting CG pair potential would be the conventional two-body pmf. Consequently, the difference between the two-body pmf, $w(r)$, and the MS-CG pair potential may be attributed entirely to the effect of three-particle correlations. Figure 3 directly compares the MS-CG force field, $f(r)$, (black curve) with the mean force, $f^{mf}(r) = -w'(r)$, (red curve) for each system. In Figure 3, panel (a) corresponds to the LJ system, while panels (b) and (c) correspond to the SPC-COG and SPC-COM systems, respectively. Figure 4 compares the MS-CG forces (panel a) and the mean forces (panel b) for the different systems. In Figure 4 solid curves describe the LJ system, dashed curves describe the SPC-COG system, and dotted curves describe the SPC-COM system. The mean force is computed from $f^{mf}(r) = k_B T g'(r) / g(r)$, where $g'(r)$ is evaluated using atomistic force information according to eq (18) and the rdf has been smoothed with a running-average over adjacent table elements. The exact LJ force field is also presented as a solid light green line in Figure 3a and it can be seen that the difference between the exact LJ force field and the MS-CG force field for the LJ system is essentially within the thickness of the lines. The mean force for both SPC models vanishes rapidly with increasing r and is relatively featureless after the first attractive well, while the mean force for the LJ model demonstrates longer-range interactions with significant repulsive and attractive regions. The degrees of freedom that have been integrated out in the coarse-grained SPC model screen the mean force between molecules. The MS-CG force field obtained from either representation of the SPC model contains two attractive wells. The first attractive well is deeper and more narrow than the attractive well in the LJ force field. The rapid decay of the mean force after the attractive well in both CG SPC models generates a repulsive barrier in the MS-CG force field at the same distance as the first minimum in the SPC rdf, which corresponds to the presence of hydrogen bonding between each SPC oxygen atom and the second nearest non-bonded hydrogen.[36] The barrier between the attractive wells is larger and occurs at shorter range for the SPC-COM model than for the SPC-COG model. The oscillations in the LJ mean force correspond to a smooth monotonic decay of the MS-CG force field.

Three-particle correlations enter the MS-CG equations through the quantity $G^{(3)}_{dd'}$, defined in eq (16), which in the continuum limit may be represented by the symmetric function

$$G^{(3)}(r,r') = \left\langle \sum_I \sum_{i} \sum_{j \neq i, k \neq i,j} \left(\vec{u}^I_{i,k} \cdot \vec{u}^I_{i,k} \right) \delta\left(r^I_{i,j} - r\right) \delta\left(r^I_{i,k} - r'\right) \right\rangle .$$

This quantity is plotted in Figure 5 as a function of r for $r' = 0.3$ and 0.6 nm . This function reflects the influence of both direct and indirect excluded volume effects on the fluid structure and, ultimately, on the MS-CG interaction potential between CG sites. The three-particle correlations couple the equations for the force field elements, f_d, according to eq (17). Figure 5 has been presented to highlight the fine structure of the three-particle correlations between CG sites within the atomistic system. The range of the coupling, though, is somewhat exaggerated by Figure 5, since the other terms in eq (17) scale as r^2 due to the statistics of particles on a shell.

For molecular scale coarse-graining (i.e., coarse-graining on length- and time- scales for which molecular excluded volume is relevant), the presence of a fixed third CG particle impacts the distribution of a given pair of CG sites. As illustrated in Figure 1, the geometry of contact between the second and third CG particles defines an excluded volume cone. This cone corresponds to a range of angles $\theta_{i;j,k} \leq \theta^*_{i;j,k} \approx 0$ that are not sampled during the MD simulation and, as a result, the spherical average of the cosine of this angle ($\cos \theta_{i;j,k} = \vec{u}_{i,j} \cdot \vec{u}_{i,k}$) does not vanish. This depletion effect is clearly evident in the calculations of Figure 5, where it can be seen that $G^{(3)}(r,r')$ has a negative peak centered at $r \approx r'$. Figure 5 also demonstrates a corresponding density enhancement resulting from the solvation shell of the third CG particle, centered roughly at $r \approx r' \pm 0.3$nm. The negative peak of the three-particle correlation function is most pronounced near the first maximum of the rdf. For larger r', the peak is more diffuse. The SPC water and LJ fluid three-particle correlations are quite similar for $r,r' > 0.3$nm and

this similarity increases with increasing r,r'. This result is perhaps quite surprising and suggests that, for simple fluids, a reasonable approximation to the three-particle correlation functions important to the MS-CG method might be obtained by appropriately rescaling the LJ three-particle correlation function. This would dramatically reduce the necessary simulation time involved in determining the MS-CG force field, since the three-particle correlation function is clearly the most difficult quantity in the MS-CG equations to converge.

4. Discussion

The analysis of section 2.1 proves that, if no approximations are made in the form of the CG force field, then the MS-CG method determines a multi-dimensional potential of mean force between CG sites according to eq (5). Simulations of CG models employing this many-body interaction potential will sample CG representations of atomistic configurations according to the *atomistic* distribution function. Such a simulation would reproduce any structural and thermodynamic properties of the atomistic system that may be observed in the CG simulation. However, the calculation and simulation of such a potential are in general not practical. Rather, CG force fields often employ pair potentials to model non-bonded interactions between CG sites. Consequently, it is important to understand the mechanism by which CG pair potentials incorporate many-body effects to approximate this multi-dimensional potential of mean force.

The YBG equation states an exact relationship between the equilibrium two- and three-particle distribution functions resulting from simulations of a given pair potential function.[20] If this pair potential depends only upon the inter-particle distance, then, for an isotropic homogeneous system, the YBG equation may be reduced to a one-dimensional integral equation that is equivalent to the MS-CG equations. Thus the MS-CG equations are equivalent in form to a statement of the exact relationship between the two- and three- particle distribution functions that arise from simulations of a homogeneous, isotropic system employing a central pair potential. (However, it should be noted that the converse of the YBG relationship does not necessarily follow. Although, a given pair potential determines a set of resulting correlation functions, a given set of correlation functions does not necessarily determine a pair potential. Thus this relationship does not directly address the general validity of approximating the many-body CG force field with a pairwise additive form.)

The MS-CG procedure may therefore be considered a novel numerical mechanism for solving the ‘inverse’ problem of liquid state theory,[19] i.e., determining an interaction pair potential that generates a given set of distribution functions. If a given set of two- and three-particle distribution functions are generated from a pair interaction, then this interaction may be determined by inverting the generalized YBG equation (26) for $\vec{F}(\vec{x}_i,\vec{x}_j)$. If, moreover, this interaction is central and depends only upon the distance between the two particles, then this pair-wise central force field may be determined by solving the MS-CG equation (23) for $f(r_{i,j})$. The MS-CG method takes advantage of atomic force information rather than directly evaluating the numerical derivative of the rdf, which may require extensive sampling.

The Ornstein-Zernicke (OZ) and YBG equations are two of the most well known integral equations for liquid state theory.[20] The ‘direct’ solution to either integral equation for the two-particle distribution function requires an (approximate) closure relation that determines a second unknown function in the integral equation. The closure for the OZ equation involves the direct correlation function, $c(r)$, while the closure relation for the YBG equation involves the three-particle distribution function. The OZ equation has been particularly useful for investigating the structure of simple liquids because the direct correlation function is short-ranged and, consequently, more amenable to theoretical and numerical analysis than the YBG equation which depends upon the three-particle correlation function, $g^{(3)}$. The OZ equation has also been employed in developing force fields for CG models.[14] However, the YBG equation

may be more useful than the OZ equation for solving the inverse problem in developing CG force fields. Although $c(r)$ may be determined by directly inverting the OZ equation, approximate closure relations are necessary to relate the direct correlation function to a CG pair potential. In contrast, the YBG equation for the CG pair potential may be closed by directly computing the relevant distributions of CG sites within an atomistic MD simulation. Inverting the resulting equation provides a force field that will reproduce the correct CG structure, under the assumption that such a force field exists.

It is a fundamental assumption that the CG distribution functions measured in atomistic MD simulations may be generated from a pair-wise central force field. In general, there may not exist such a CG pair potential that will reproduce both the two- and three- particle distribution functions characterizing a given system. Consequently the MS-CG equations presented in eq (21) are not necessarily an exact identity describing the CG structure because the assumed pair potential giving rise to the measured CG two- and three-particle distribution functions may not exist. In fact, recent work has demonstrated that many-body[37–41] and non-central[42,43] interactions may be critical for the coarse-grained modeling of proteins.

The underlying variational principle provides tremendous flexibility in the MS-CG methodology. By minimizing the residual in eq (1), the MS-CG method systematically determines an optimal approximation to the multi-dimensional many-body CG PMF. If no assumptions are made in the form of the CG interaction, the MS-CG procedure determines this many-body PMF, which would reproduce both two- and three-particle structural correlations in CG simulations. Previous numerical implementations of the MS-CG procedure[16,17,24–28] have determined and employed an optimal central pair decomposition of this CG PMF. However, in principle, the MS-CG procedure may be readily generalized to incorporate both non-central and many-body effects within the CG force field by relaxing the assumptions explicit in eq (7)–eq (8). Many-body interactions may be incorporated into the MS-CG force field either by generalizing the approach of eq (9) and tabulating additional many-body interaction terms on a multidimensional grid or by parameterizing an assumed functional form by minimizing the MS-CG residual. Similarly, non-central interactions may also be incorporated into the CG force field, for instance, by expanding the pair interaction in eq (8) as a series of spherical harmonics.[43] However, if the interaction does not depend linearly upon the force field parameters, the minimization procedure for optimizing the MS-CG force field requires a nonlinear regression algorithm.[22]

The reverse Monte Carlo (RMC) method has been employed to determine a CG potential that reproduces a given CG rdf when used in CG simulations.[2,29] Chayes and Chayes[19] have proven that there does indeed exist a unique[44] pair potential that will reproduce an observed rdf.[14] However, this potential is not guaranteed to reproduce higher order distribution functions.[31] Although by design the RMC method should reproduce CG rdf's to the desired numerical precision,[2,29] simulations employing this pair potential may not necessarily reproduce the correct three-particle correlations for the system. In particular, using a similar RMC procedure to generate water configurations from known radial distribution functions, Jedlovsky et al.[45] have demonstrated that three-particle correlations may be inaccurately reproduced although the observed rdf's are quantitatively accurate. In contrast, Iuchi et al.[46] have demonstrated that the MS-CG procedure accurately reproduces both the two- and three-particle correlations of a polarizable 4-site water model using a non-polarizable 4-site MS-CG force field.

It is of some interest to briefly compare the MS-CG and RMC methods for developing CG potentials in light of the YBG equation for homogeneous isotropic systems. The MS-CG method implicitly measures the two- and three- particle correlation functions describing CG sites within an atomistic MD simulation and then directly inverts the YBG equation to

determine a central CG pair potential that would generate these distribution functions, should such a potential exist. In general such a potential may not exist and, consequently, simulations employing the MS-CG pair potential will not necessarily reproduce either the two- or three-particle correlations exactly. Rather CG simulations employing the MS-CG force field will satisfy a different YBG equation relating the fixed MS-CG force field and the resulting CG distribution functions. Although the MS-CG force field does not necessarily reproduce either the two- or three-particle distribution functions exactly, the method clearly incorporates three-particle correlations. Moreover, because the method does not necessarily reproduce the pair correlation functions, comparison of the relevant rdf's between atomistic and CG MD simulations is a useful measure of the validity of the CG model.

In contrast, the RMC method only directly considers the two-particle correlation function and attempts to solve the YBG equation for the pair interaction that reproduces the target rdf while allowing the three-particle correlation function to vary as necessary.[2,29] The repeated MC or MD simulations used in iteratively updating the pair potential may be considered a nonlinear regression algorithm that tries to solve the YBG equation for a pair potential reproducing a fixed rdf. If three-particle correlations were not significant in the YBG equation, then the pair potential would be simply the two-body pmf, which is often implemented as an initial condition in the search for the optimal pair potential. With succesive iterations the simulated rdf converges to the target rdf measured from atomistic MD simulations.[2] The RMC method implicitly incorporates information regarding three-particle correlations by updating the force field to improve agreement between the measured and target rdf, although in successive simulations the three-particle correlations may change. The YBG equation that is implicitly solved through the RMC method incorporates the target rdf and is guaranteed to reproduce this rdf. However, the three-particle correlations in the final YBG equation may be different than those in the original atomistic representation of the system.

In closing, two additional points require brief discussion. The preceding analysis directly addresses non-bonded interactions between CG sites. Bonded interactions may be treated as special cases of non-bonded interactions or by introducing additional terms into the MS-CG potential. In the latter case, the contributions from these interactions should be subtracted from the total forces on CG sites and the preceding analysis still holds for non-bonded interactions. Additionally, it has been mentioned repeatedly that the MS-CG equations are equivalent to the YBG equation for a homogeneous, isotropic system evolving under a central pair potential. In principle, a system described by a central pair potential may be both homogeneous and isotropic. However, these symmetries are not necessarily realized in MD simulations of complex interfacial or biological systems. The presence of a lipid bilayer or a protein in an MD simulation box may break this symmetry. The analysis of Section 2.3, though, requires the additional assumptions of isotropy and homogeneity. Consequently, the FM equations are rigorously equivalent to the YBG equations in eq (35) only for relatively simple systems. However, the analysis of Section 2.2 demonstrates that the FM force field incorporates critical information regarding three-particle correlations for any system. It has been empirically demonstrated that even for highly complex systems the MS-CG method generates a useful and quantitatively accurate model.[16,17,24–28,46]

5. Conclusions

The present work develops a statistical mechanical framework for understanding the foundations and success of the MS-CG method. It has been shown that the MS-CG method may be used to derive a multi-dimensional pmf for the interactions between CG sites. The "normal" MS-CG equations have been derived for approximating this many-body interaction with central pair potentials. The derivation demonstrates that the MS-CG equations for these potentials reflect not only two-body, but also three-body correlations observed between CG

sites during atomistic MD simulations. The generalized YBG equation has been presented for CG systems and it has been proven that this generalized YBG equation is equivalent to the normal MS-CG equations for a homogeneous, isotropic system evolving under a central pair potential. The present analysis provides a connection between the MS-CG method and equilibrium statistical mechanics and also illuminates the general significance of three-particle correlations for deriving CG effective pair potentials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

W.G. Noid acknowledges funding from the NIH through a Ruth L. Kirschstein NRSA postdoctoral fellowship grant number 5F32GM076839 - 02. This research was also supported in part by the National Science Foundation (CHE-0628257). An allocation of computer time from the Center for High Performance Computing at the University of Utah is gratefully acknowledged. The computational resources for this project have been provided by the National Institutes of Health (Grant # NCRR 1 S10 RR17214-01) on the Arches Metacluster, administered by the University of Utah Center for High Performance Computing. W.G. Noid acknowledges many stimulating conversations with Dr. V. Krishna.

Appendices

Appendices A and B involve the properties of Dirac delta functions in spherical polar coordinates. In particular, the derivation of eq (18) and eq (33) employ the following relation[47]

$$\int d\Omega A(\vec{r})\delta^{(3)}(\vec{r}-\vec{r}_0)=\frac{1}{r^2}A_{SP}(r,\Omega_0)\delta(r-r_0), \tag{36}$$

which is valid for all $r_0 \neq 0$. In eq (36), $A_{SP}(r,\Omega_0)$ is the representation of the function $A(\vec{r})$ in spherical polar coordinates evaluated in the direction Ω_0 defined by the vector $\vec{r}_0$.

Appendix A. Derivation of Equation (18)

Assume that the CG configurations employed in the FM procedure were sampled according to a distribution function, $\rho_{CG}\left(\vec{r}^{N_{CG}}\right)=\exp\left[-V^{CG}\left(\vec{r}^{N_{CG}}\right)/k_BT\right]/Z(N,V_0,T)$, where $V^{CG}\left(\vec{r}^{N_{CG}}\right)$ is a many-particle CG potential function describing the interactions between CG sites. Selecting two arbitrary CG sites, i and j, and transforming into "sum," $\vec{R}_{ij}=\frac{1}{2}\left(\vec{r}_i+\vec{r}_j\right)$, and "difference," $\vec{r}_{ij}=\vec{r}_i-\vec{r}_j$, variables, then it follows that:

$$\left(\frac{\partial\rho_{CG}\left(\vec{r}^{N_{CG}}\right)}{\partial r_{ij}}\right)_{\vec{R}_{ij},\Omega_{ij},\vec{r}^{(N_{CG}-2)}}=\frac{1}{2k_BT}\left(\left(\vec{F}_i-\vec{F}_j\right)\cdot\vec{u}_{ij}\right)\rho_{CG}\left(\vec{r}^{N_{CG}}\right). \tag{37}$$

The partial derivative is evaluated with respect to the inter-CG site distance (difference variable) r_{ij}, while holding fixed the relative (difference) orientation Ω_{ij}, and average position of the two CG sites $\vec{R}_{ij}$ as well as the remaining $N_{CG}-2$ sites, $\vec{r}^{(N_{CG}-2)}$.

Consider next the quantity $G_{ij}(r) = \langle \delta(r - r_{ij}) \rangle$. Employing the identity in eq (36), it can be proven that

$$\frac{d}{dr} G_{ij}(r) = \frac{2}{r} G_{ij}(r) + \frac{1}{2k_B T} \langle \left(\vec{F}_i - \vec{F}_j \right) \cdot \vec{u}_{ij} \delta(r - r_{ij}) \rangle. \tag{38}$$

Let i and j correspond to CG site types α and β, respectively, and define the radial distribution function

$$g_{\alpha\beta}(r) = \left(\frac{1}{4\pi \rho_\alpha \rho_\beta V_0} \right) \frac{1}{r^2} \langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha} G_{i_\alpha j_\beta}(r) \rangle, \tag{39}$$

then

$$\frac{d}{dr} g_{\alpha\beta}(r) = \frac{1}{2k_B T} \left(\frac{1}{4\pi \rho_\alpha \rho_\beta V_0} \right) \frac{1}{r^2} \langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha} \left(\vec{F}_{i_\alpha} - \vec{F}_{j_\beta} \right) \cdot \vec{u}_{i_\alpha j_\beta} \delta(r - r_{i_\alpha j_\beta}) \rangle. \tag{40}$$

In the case that α = β, the two terms in the summations are equivalent and the result simplifies:

$$\frac{d}{dr} g_{\alpha\alpha}(r) = \frac{1}{k_B T} \left(\frac{1}{4\pi \rho_\alpha \rho_\alpha V_0} \right) \frac{1}{r^2} \langle \sum_{i_\alpha} \sum_{j_\alpha \neq i_\alpha} \left(\vec{F}_{i_\alpha} \cdot \vec{u}_{i_\alpha j_\alpha} \right) \delta(r - r_{i_\alpha j_\alpha}) \rangle. \tag{41}$$

A more detailed derivation is provided in the supporting information section.

Appendix B. Derivation of Equation (33)

Consider the distribution of three distinct CG particles at positions $\vec{r}_1$, $\vec{r}_2$, and $\vec{r}_3$ corresponding to particles i_α, j_β, and $k\gamma$ in Figure 1. The FM kernel is defined

$$M^{FM}_{\alpha;\beta\gamma}(r_{12}, r_{13}) = \left(\frac{1}{4\pi \rho_\alpha \rho_\beta V_0} \right) \frac{1}{{r_{12}}^2} \langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha, k_\gamma} \sum_{k_\gamma \neq i_\alpha, j_\beta} \left(\vec{u}^I_{i_\alpha, j_\beta} \cdot \vec{u}^I_{i_\alpha, k_\gamma} \right) \delta_M(r^I_{i_\alpha, j_\beta} - r_{12}) \delta_M \left(r^I_{i_\alpha, k_\gamma} - r_{13} \right) \rangle_I \tag{42}$$

The YBG kernel is defined:

$$M^{YBG}_{\alpha;\beta\gamma}(r_{12}, r_{13}) = \rho_\gamma {r_{13}}^2 \int d\Omega_{31} \int \frac{1}{4\pi} d\Omega_{21} \int \frac{1}{V_0} d\vec{r}_1 (\vec{u}_{12} \cdot \vec{u}_{13}) g^{(3)}_{\alpha\beta\gamma}(\vec{r}_1, \vec{r}_2, \vec{r}_3) \tag{43}$$

where Ω_{n1} defines the orientation of the unit vector $\vec{u}_n$ and the three-particle correlation function is defined:

$$g^{(3)}_{\alpha\beta\gamma}\left(\vec{r}_1, \vec{r}_2, \vec{r}_3 \right) = \frac{1}{\rho_\alpha \rho_\beta \rho_\gamma} \langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha, k_\gamma} \sum_{k_\gamma \neq i_\alpha, j_\beta} \delta^{(3)} \left(\vec{r}^I_{i_\alpha} - \vec{r}_1 \right) \delta^{(3)} \left(\vec{r}^I_{j_\beta} - \vec{r}_2 \right) \delta^{(3)} \left(\vec{r}^I_{k_\gamma} - \vec{r}_3 \right) \rangle_I . \tag{44}$$

The definitions in eq (42) and eq (44) assume that the average over configurations is equivalent to the appropriate ensemble average. After integrating over $\vec{r}_1$ in eq (43), the YBG kernel may be expressed

$$M^{YGG}_{\alpha;\beta\gamma}(r_{12},r_{13}) = \frac{r_{13}^2}{4\pi\rho_\alpha\rho_\beta v_0}\Big\langle \sum_{i_\alpha}\sum_{j_\beta\neq i_\alpha,\kappa_\gamma\kappa_\gamma\neq i_\alpha,j_\beta}\int d\Omega_{31}\int d\Omega_{21}\left(\vec{u}_{12}(\Omega_{21})\cdot\vec{u}_{13}(\Omega_{31})\right)\delta^{(3)}\left(\vec{r}^{\,I}_{j_\beta,i_\alpha}-\vec{r}_{21}\right)\delta^{(3)}\left(\vec{r}^{\,I}_{\kappa_\gamma,i_\alpha}-\vec{r}_{31}\right)\Big\rangle_I \tag{45}$$

The integrals over the orientational degrees of freedom may then be evaluated according to the identity in eq (36). Upon performing these integrals within the ensemble average, one obtains the desired result:

$$\begin{aligned} M^{YBG}_{\alpha;\beta\gamma}(r_{12},r_{13}) &= \left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r_{12}^2}\Big\langle \sum_{i_\alpha}\sum_{j_\beta\neq i_\alpha,k_\gamma k_\gamma\neq i_\alpha,j_\beta}\left(\vec{u}^{\,I}_{i_\alpha,j_\beta}\cdot\vec{u}^{\,I}_{i_\alpha,k_\gamma}\right)\delta\left(r^I_{i_a,j_\beta}-r_{12}\right)\delta(r^I_{i_\alpha,k_\gamma}-r_{13})\Big\rangle_I \\ &= M^{FM}_{\alpha;\beta\gamma}(r_{12},r_{13}) \end{aligned} \tag{46}$$

Appendix C. Normal FM equations for a multi-component system

The FM residual for a system with multiple types of CG sites may be expressed:

$$\chi^2=\frac{1}{3N_I N_{CG}}\sum_I^{N_I}\sum_\alpha^{N_T}\sum_{i_\alpha}^{N_\alpha}\left|\vec{F}^{\,I,AA}_{i_\alpha}-\vec{F}^{\,I,CG}_{i_\alpha}\right|^2. \tag{47}$$

In eq (47) and in the following, Greek indices (α) represent types of CG sites, Latin subscripts (*i*) represent particular sites of a given CG type, and the superscript (*I*) labels configurations sampled during the atomistic MD simulation. As before it is assumed that the FM force field is pair-wise additive, such that

$$\vec{F}^{\,I,CG}_{i_\alpha}=\sum_\beta\sum_{j_\beta\neq i_\alpha}\vec{F}^{\,CG}_{i_\alpha,j_\beta}\left(\vec{r}^{\,I}_{i_\alpha},\vec{r}^{\,I}_{j_\beta}\right) \tag{48}$$

and, furthermore, that this force field is central:

$$\vec{F}^{\,CG}_{i_\alpha,j_\beta}\left(\vec{r}^{\,I}_{i_\alpha},\vec{r}^{\,I}_{j_\beta}\right)=\vec{u}^{\,I}_{i_\alpha,j_\beta}f^{\alpha\beta}\left(\vec{r}^{\,I}_{i_\alpha,j_\beta}\right). \tag{49}$$

Employing the discrete delta function basis of eq (9), a system of linear equations (i.e., the "normal" FM equations for a multi-component system) is obtained by minimizing the residual in eq (47) with respect to each interaction, $f_d^{\alpha\beta}$:

$$\sum_\gamma\sum_{d'}\left[\left(1-\delta_{\alpha\beta}\right)f^{\alpha\gamma}_{d'}G^{\alpha;\beta\gamma}_{dd'}+f^{\beta\gamma}_{d'}G^{\beta;\alpha\gamma}_{dd'}\right]=\left(1-\delta_{\alpha\beta}\right)b^{\alpha\beta}_d+b^{\beta\alpha}_d, \tag{50}$$

in which

$$b_d^{\alpha\beta}=\left\langle \sum_{i_\alpha}\sum_{j_\beta\neq i_\alpha}\left(\vec{F}_{i_\alpha}^{I,AA}\cdot\vec{u}_{i_\alpha,j_\beta}^{I}\right)\delta_D\left(r_{i_\alpha,j_\beta}^{I}-r_d\right)\right\rangle_I , \tag{51}$$

and

$$G_{dd'}^{\alpha;\beta\gamma}=\left\langle \sum_{i_\alpha}\sum_{j_\beta\neq i_\alpha}\sum_{k_\gamma\neq i_\alpha}\left(\vec{u}_{i_\alpha,j_\beta}^{I}\cdot\vec{u}_{i_\alpha,k_\gamma}^{I}\right)\delta_D\left(r_{i_\alpha,j_\beta}^{I}-r_d\right)\delta_D\left(r_{i_\alpha,k_\gamma}^{I}-r_{d'}\right)\right\rangle_I . \tag{52}$$

Equation (50) generalizes eq (10) and employs the symmetry $f_d^{\alpha\beta}=f_d^{\beta\alpha}$. The quantity defined in eq (52) may be decomposed into two- and three-particle contributions according to:

$$G_{dd'}^{\alpha;\beta\gamma}=\delta_{\beta\gamma}\delta_{dd'}\left(G^{(2)}\right)_d^{\alpha\beta}+\left(G^{(3)}\right)_{dd'}^{\alpha;\beta\gamma}, \tag{53}$$

where

$$\left(G^{(2)}\right)_d^{\alpha\beta}=\left\langle \sum_{i_\alpha}\sum_{j_\beta\neq i_\alpha}\delta_D\left(r_{i_\alpha,j_\beta}^{I}-r_d\right)\right\rangle_I \tag{54}$$

and

$$\left(G^{(3)}\right)_{dd'}^{\alpha;\beta\gamma}=\left\langle \sum_{i_\alpha}\sum_{j_\beta\neq i_\alpha,k_\gamma}\sum_{k_\gamma\neq i_\alpha,j_\beta}\left(\vec{u}_{i_\alpha,j_\beta}^{I}\cdot\vec{u}_{i_\alpha,k_\gamma}^{I}\right)\delta_D\left(r_{i_\alpha,j_\beta}^{I}-r_d\right)\delta_D\left(r_{i_\alpha,k_\gamma}^{I}-r_{d'}\right)\right\rangle_I . \tag{55}$$

Employing these definitions, eq (50) may then be re-expressed:

$$\begin{aligned}\left(1-\delta_{\alpha\beta}\right)\left[b_d^{\alpha\beta}-f_d^{\alpha\beta}\left(G^{(2)}\right)_d^{\alpha\beta}\right]+\left[b_d^{\beta\alpha}-f_d^{\alpha\beta}\left(G^{(2)}\right)_d^{\beta\alpha}\right]=\\ \sum_\gamma\sum_{d'}\left[\left(1-\delta_{\alpha\beta}\right)f_{d'}^{\alpha\gamma}\left(G^{(3)}\right)_{dd'}^{\alpha;\beta\gamma}f_{d'}^{\beta\gamma}\left(G^{(3)}\right)_{dd'}^{\beta;\alpha\gamma}\right]\end{aligned} \tag{56}$$

Upon multiplication by the factor $1/(4\pi\rho_\alpha\rho_\beta V_0\cdot r_d^2)$ and application of the definition in eq (39) and the identity in eq (41) from Appendix A, eq (56) may be expressed:

$$k_BT\frac{dg_{\alpha\beta}}{dr}\Big|_d-f_d^{\alpha\beta}g_{\alpha\beta}\Big|_d=\sum_\gamma\sum_{d'}\frac{1}{2}\left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r_d^2}\left[f_{d'}^{\alpha\gamma}\left(G^{(3)}\right)_{dd'}^{\alpha;\beta\gamma}+f_{d'}^{\beta\gamma}\left(G^{(3)}\right)_{dd'}^{\beta;\alpha\gamma}\right]. \tag{57}$$

Passing into the continuum limit, eq (57) may be represented by a linear one-dimensional integral equation:

$$\left(k_BT\frac{d}{dr}-f^{\alpha\beta}(r)\right)g_{\alpha\beta}(r)=\sum_\gamma\int dr'\,\frac{1}{2}\left[f^{\alpha\gamma}(r')\,M_{\alpha;\beta\gamma}^{FM}(r,r')+f^{\beta\gamma}(r')\,M_{\beta;\alpha\gamma}^{FM}(r,r')\right], \tag{58}$$

where

$$M^{FM}_{\alpha;\beta\gamma}(r,r') = \left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\langle\sum_{i_\alpha}\sum_{j_\beta\neq i_\alpha,k_\gamma}\sum_{k_\gamma\neq i_\alpha,j_\beta}\left(\vec{u}^{\,I}_{i_\alpha,j_\beta}\cdot\vec{u}^{\,I}_{i_\alpha,k_\gamma}\right)\delta\left(r^I_{i_\alpha,j_\beta}-r\right)\delta\left(r^I_{i_\alpha,k_\gamma}-r'\right)\rangle_I , \tag{59}$$

and, according to Appendix B, $M^{FM}_{\alpha;\beta\gamma}(r,r') = M^{YBG}_{\alpha;\beta\gamma}(r,r') = M_{\alpha;\beta\gamma}(r,r')$.

Appendix D. Generalized YBG equation for multi-component CG system

The Hamiltonian for a system with N_α sites of type α and $N_{CG}=\sum_\alpha^{N_T} N_\alpha$ total CG sites evolving under an N_{CG} particle potential function, $V^{N_{CG}}$, is defined:

$$H\left(\vec{p}^{\,N_{CG}},\vec{r}^{\,N_{CG}}\right)=\sum_\alpha^{N_T}\sum_{i_\alpha}^{N_\alpha}\frac{1}{2m_\alpha}\left|\vec{p}_{i_\alpha}\right|^2+V^{N_{CG}}\left(\vec{r}^{\,N_{CG}}\right). \tag{60}$$

Assuming that the total force on each CG site, $\vec{F}_{i_\alpha}=-\partial V^{N_{CG}}/\partial\vec{r}_{i_\alpha}$, arises from a sum of pair interactions,

$$\vec{F}_{i_\alpha}=\sum_\beta\sum_{j_\beta}\vec{F}_{i_\alpha,j_\beta}\left(\vec{r}_{i_\alpha},\vec{r}_{j_\beta}\right), \tag{61}$$

then for a homogeneous system with $\rho_\alpha(\vec{r}_{i_\alpha})=\rho_\alpha=const$, the following relations describe the distribution of three arbitrary but distinct CG particles, i_α, j_β, and $k\gamma$:

$$\begin{aligned}\left(k_BT\frac{\partial}{\partial\vec{r}_{i_\alpha}}-\vec{F}_{i_\alpha,j_\beta}\right)g^{(2)}_{\alpha\beta}\left(\vec{r}_{i_\alpha},\vec{r}_{j_\beta}\right)&=\sum_\gamma\rho_\gamma\int d\vec{r}_{k_\gamma}\vec{F}_{i_\alpha,k_\gamma}g^{(3)}_{\alpha\beta\gamma}\left(\vec{r}_{i_\alpha},\vec{r}_{j_\beta},\vec{r}_{k_\gamma}\right)\\ \left(k_BT\frac{\partial}{\partial\vec{r}_{j_\beta}}-\vec{F}_{j_\beta,i_\alpha}\right)g^{(2)}_{\alpha\beta}\left(\vec{r}_{i_\alpha},\vec{r}_{j_\beta}\right)&=\sum_\gamma\rho_\gamma\int d\vec{r}_{k_\gamma}\vec{F}_{j_\beta,k_\gamma}g^{(3)}_{\alpha\beta\gamma}\left(\vec{r}_{i_\alpha},\vec{r}_{j_\beta},\vec{r}_{k_\gamma}\right)\end{aligned} . \tag{62}$$

Equations (62) generalize the Yvon-Born-Green (YBG) equation[20] and relate the equilibrium two- and three- CG particle distribution functions to the pair-wise decomposable force field, $\vec{F}_{i_\alpha,j_\beta}$. As before, for an isotropic homogeneous system, the two-particle distribution function is equal to its rotational and translational average, $g^{(2)}_{\alpha\beta}\left(\vec{r}_{i_\alpha},\vec{r}_{j_\beta}\right)=g_{\alpha\beta}\left(r_{i_\alpha,j_\beta}\right)$, and the gradient terms in eqs (62) may then be simplified: $\partial g^{(2)}_{\alpha\beta}\left(\vec{r}_{i_\alpha},\vec{r}_{j_\beta}\right)/\partial\vec{r}_{i_\alpha}=\vec{u}_{i_\alpha,j_\beta}\,dg_{\alpha\beta}\left(r_{i_\alpha,j_\beta}\right)/dr_{i_\alpha,j_\beta}$. Under the additional assumption that the pair interaction between CG sites is central, $\vec{F}_{i_\alpha,j_\beta}\left(\vec{r}_{i_\alpha},\vec{r}_{j_\beta}\right)=\vec{u}_{i_\alpha,j_\beta}f^{\alpha\beta}\left(r_{i_\alpha,j_\beta}\right)$, the first equality in eq (62) may be re-expressed:

$$\vec{u}_{i_\alpha,j_\beta}\left(k_BT\frac{d}{dr_{i_\alpha,j_\beta}}-f^{\alpha\beta}\left(r_{i_\alpha,j_\beta}\right)\right)g_{\alpha\beta}\left(r_{i_\alpha,j_\beta}\right)=\sum_\gamma\rho_\gamma\int d\vec{r}_{k_\gamma}\vec{u}_{i_\alpha,k_\gamma}f^{\alpha\gamma}\left(r_{i_\alpha,k_\gamma}\right)g^{(3)}_{\alpha\beta\gamma}\left(\vec{r}_{i_\alpha},\vec{r}_{j_\beta},\vec{r}_{k_\gamma}\right). \tag{63}$$

Projecting this equation onto the vector $\vec{u}_{i\alpha,j\beta}$ and shifting the integration variable, one obtains the result:

$$\left(k_B T \frac{d}{dr_{i_\alpha,j_\beta}} - f^{\alpha\beta}\left(r_{i_\alpha,j_\beta}\right)\right) g_{\alpha\beta}\left(r_{i_\alpha,j_\beta}\right) = \sum_\gamma \int dr_{i_\alpha,k_\gamma} f^{\alpha\gamma}\left(r_{i_\alpha,k_\gamma}\right) M^{YBG}_{\alpha;\beta\gamma}\left(r_{i_\alpha,j_\beta}, r_{i_\alpha,k_\gamma}\right), \tag{64}$$

where,

$$M^{YBG}_{\alpha;\beta\gamma}\left(r_{i_\alpha,j_\beta}, r_{i_\alpha,k_\gamma}\right) = \rho_\gamma\left(r_{i_\alpha,k_\gamma}\right)^2 \int d\Omega_{k_\gamma,i_\alpha} \hat{P}_{\mathrm{RT}}\left(\vec{u}_{i_\alpha,k_\gamma} \cdot \vec{u}_{i_\alpha,j_\beta}\right) g^{(3)}_{\alpha\beta\gamma}\left(\vec{r}_{i_\alpha}, \vec{r}_{j_\beta}, \vec{r}_{k_\gamma}\right) \tag{65}$$

and, according to Appendix B, $M^{YBG}_{\alpha;\beta\gamma}(r,r') = M^{FM}_{\alpha;\beta\gamma}(r,r') = M_{\alpha;\beta\gamma}(r,r')$. Similar manipulation of the second relation in eq (62) yields

$$\left(k_B T \frac{d}{dr_{i_\alpha,j_\beta}} - f^{\alpha\beta}\left(r_{i_\alpha,j_\beta}\right)\right) g_{\alpha\beta}\left(r_{i_\alpha,j_\beta}\right) = \sum_\gamma \int dr_{j_\beta,k_\gamma} f^{\beta\gamma}\left(r_{j_\beta,k_\gamma}\right) M_{\beta;\alpha\gamma}\left(r_{i_\alpha,j_\beta}, r_{j_\alpha,k_\gamma}\right). \tag{66}$$

The left hand sides and thus the right hand sides of eq (64) and eq (66) are identical. The two equations may then be averaged to yield the final result, the YBG equation for a homogeneous, isotropic multi-component CG system evolving under a central pair potential:

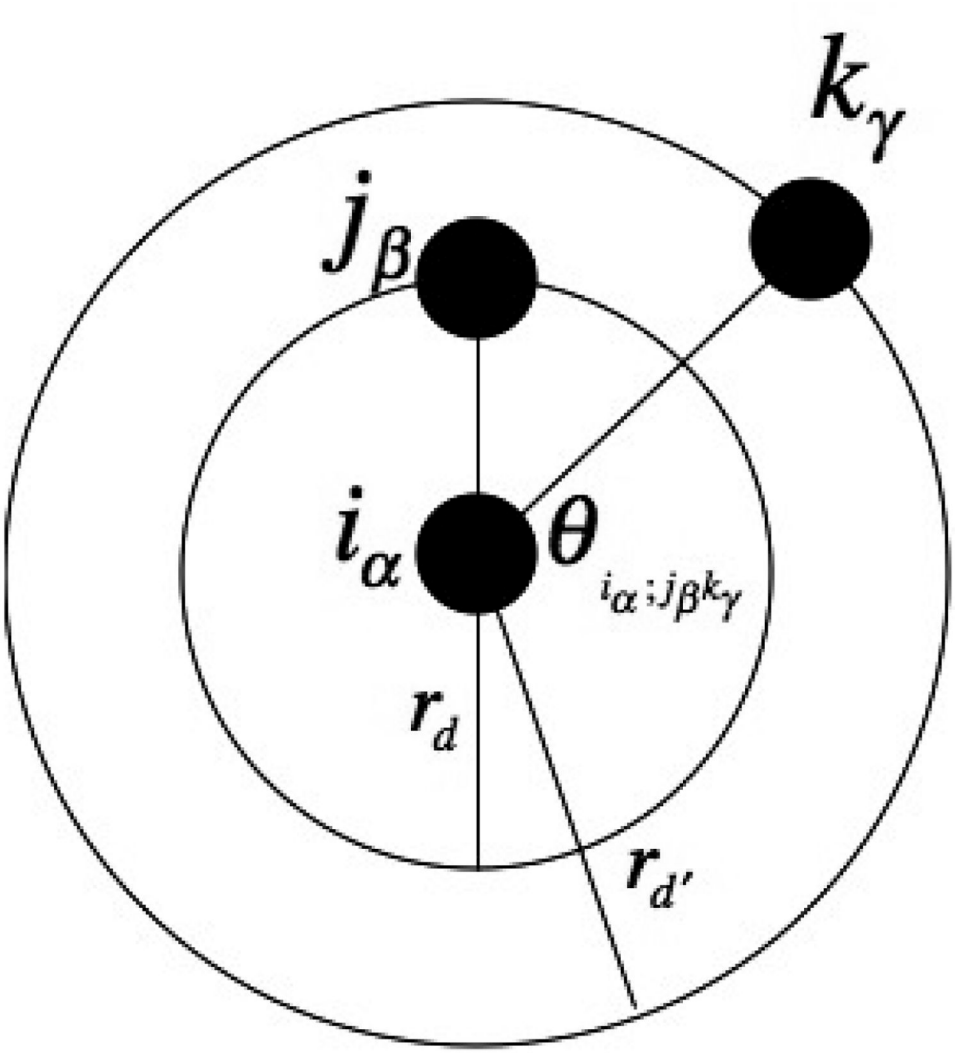
$$\left(k_B T \frac{d}{dr} - f^{\alpha\beta}(r)\right) g_{\alpha\beta}(r) = \sum_\gamma \int dr' \tfrac{1}{2}\left[f^{\alpha\gamma}(r') M_{\alpha;\beta\gamma}(r,r') + f^{\beta\gamma}(r') M_{\beta;\alpha\gamma}(r,r')\right]. \tag{67}$$

Therefore, the generalized YBG equation for a homogeneous and isotropic multi-component CG system evolving under a central pair potential is equivalent to the normal multi-component FM eqs (58).

References

1. Karplus M, McCammon JA. Nat. Struct. Biol 2002;9:646. [PubMed: 12198485]
2. Muller-Plathe F. Chemphyschem 2002;3:754.
3. Duan Y, Kollman PA. Science 1998;282:740. [PubMed: 9784131]
4. Daggett V. Chem. Rev 2006;106:1898. [PubMed: 16683760]
5. Wulfing C, Sjaastad MD, Davis MM. Proc. Natl. Acad. Sci. U. S. A 1998;95:6302. [PubMed: 9600960]
6. Akkermans RLC, Briels WJ. J. Chem. Phys 2000;113:6409.
7. Akkermans RLC, Briels WJ. J. Chem. Phys 2001;114:1020.
8. Forrest BM, Suter UW. J. Chem. Phys 1995;102:7256.
9. Shelley JC, Shelley MY, Reeder RC, Bandyopadhyay S, Klein ML. J. Phys. Chem. B 2001;105:4464.
10. Lopez CF, Moore PB, Shelley JC, Shelley MY, Klein ML. Comput. Phys. Commun 2002;147:1.
11. Nielsen SO, Lopez CF, Ivanov I, Moore PB, Shelley JC, Klein ML. Biophys. J 2004;87:2107. [PubMed: 15454415]
12. Nielsen SO, Lopez CF, Srinivas G, Klein ML. J. Phys.: Condens. Matter 2004;16:R481.
13. Bolhuis PG, Louis AA, Hansen JP. Phys. Rev. E 2001;6402
14. Louis AA, Bolhuis PG, Hansen JP, Meijer EJ. Phys. Rev. Lett 2000;85:2522. [PubMed: 10978097]
15. Marrink SJ, de Vries AH, Mark AE. J. Phys. Chem. B 2004;108:750.
16. Izvekov S, Voth GA. J. Chem. Phys 2005;123:134105. [PubMed: 16223273]

17. Izvekov S, Voth GA. J. Phys. Chem. B 2005;109:2469. [PubMed: 16851243]
18. See, for example, the recent work inJ. Chem. Theory Comput 2006;2(3)and references cited therein.
19. Chayes JT, Chayes L. J. Stat. Phys 1984;36:471.
20. Hansen, JP.; McDonald, IR. Theory of Simple Liquids. Vol. 2 ed. San Diego, CA: Academic Press; 1990.
21. Reatto L, Levesque D, Weis JJ. Phys. Rev. A 1986;33:3451. [PubMed: 9897057]
22. Ercolessi F, Adams JB. Europhys. Lett 1994;26:583.
23. Izvekov S, Parrinello M, Burnham CJ, Voth GA. J. Chem. Phys 2004;120:10896. [PubMed: 15268120]
24. Wang YT, Izvekov S, Yan TY, Voth GA. J. Phys. Chem. B 2006;110:3564. [PubMed: 16494412]
25. Izvekov S, Voth GA. J. Chem. Theory and Comput 2006;2:637.
26. Zhou J, Thorpe IF, Izvekov S, Voth GA. Biophys. J. 2007in press
27. Izvekov S, Violi A, Voth GA. J. Phys. Chem. B 2005;109:17019. [PubMed: 16853168]
28. Shi Q, Izvekov S, Voth GA. J. Phys. Chem. B 2006;110:15045. [PubMed: 16884212]
29. Lyubartsev A, Laaksonen A. Phys. Rev. E 1995;52:3730.
30. Schulman, LS. Techniques and applications of path integration. John Wiley and Sons; 1981.
31. Evans R. Mol. Sim 1990;4:409.
32. Berendsen, HJ.; Postma, JPM.; van Gunsteren, WF.; Hermans, J. Intermolecular Forces. Pullman, B., editor. Reidel, Dordrecht; 1981. p. 331
33. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. J. Comput. Chem 2005;26:1701. [PubMed: 16211538]
34. Hoover WG. Phys. Rev. A 1985;31:1695. [PubMed: 9895674]
35. Nose S. Mol. Phys 1984;52:255.
36. Chandler, D. Introduction to modern statistical mechanics. Oxford University Press; 1987.
37. Kolinski A, Skolnick J. J. Chem. Phys 1992;97:9412.
38. Kolinski A, Skolnick J. Proteins: Structure, Function, and Genetics 1994;18:338.
39. Liwo A, Czaplewski C, Pillardy J, Scheraga HA. J. Chem. Phys 2001;115:2323.
40. Liwo A, Kazmierkiewicz R, Czaplewski C, Groth M, Oldziej S, Wawak RJ, Rackovsky S, Pincus MR, Scheraga HA. Journal of Computational Chemistry 1998;19:259.
41. Vendruscolo M, Domany E. J. Chem. Phys 1998;109:11101.
42. Buchete NV, Straub JE. J. Chem. Phys 2003;118:7658.
43. Buchete NV, Straub JE, Thirumalai D. Journal of Molecular Graphics and Modeling 2004;22:441.
44. Henderson RL. Phys. Lett 1974;49A:197.
45. Jedlovszky P, Bako I, Palinkas G, Radnai T, Soper AK. J. Chem. Phys 1996;105:245.
46. Iuchi S, Izvekov S, Voth GA. J. Chem. Phys. 2007in press
47. Arfken, GB.; Weber, HJ. Mathematical methods for physicists. San Diego: Academic Press; 1995.



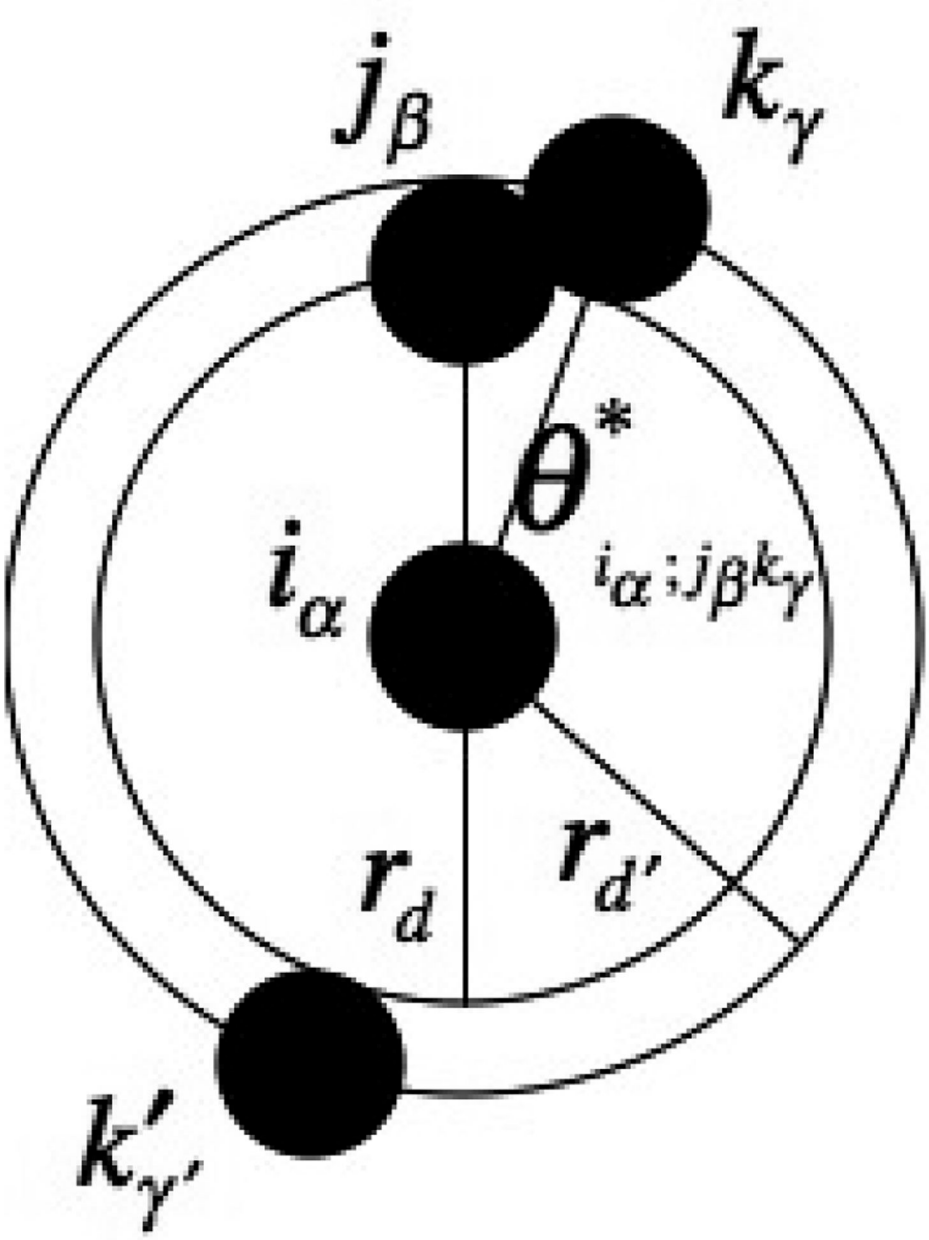


Figure 1.

(a) The geometry of three-particle correlations between CG sites relevant to the MS-CG equations is schematically illustrated. (b) Excluded volume effects define a cone of small angles $\theta_{i_\alpha;j_\beta k_\gamma} \leq \theta^*_{i_\alpha;j_\beta k_\gamma}$ that is not sampled during MD simulations.

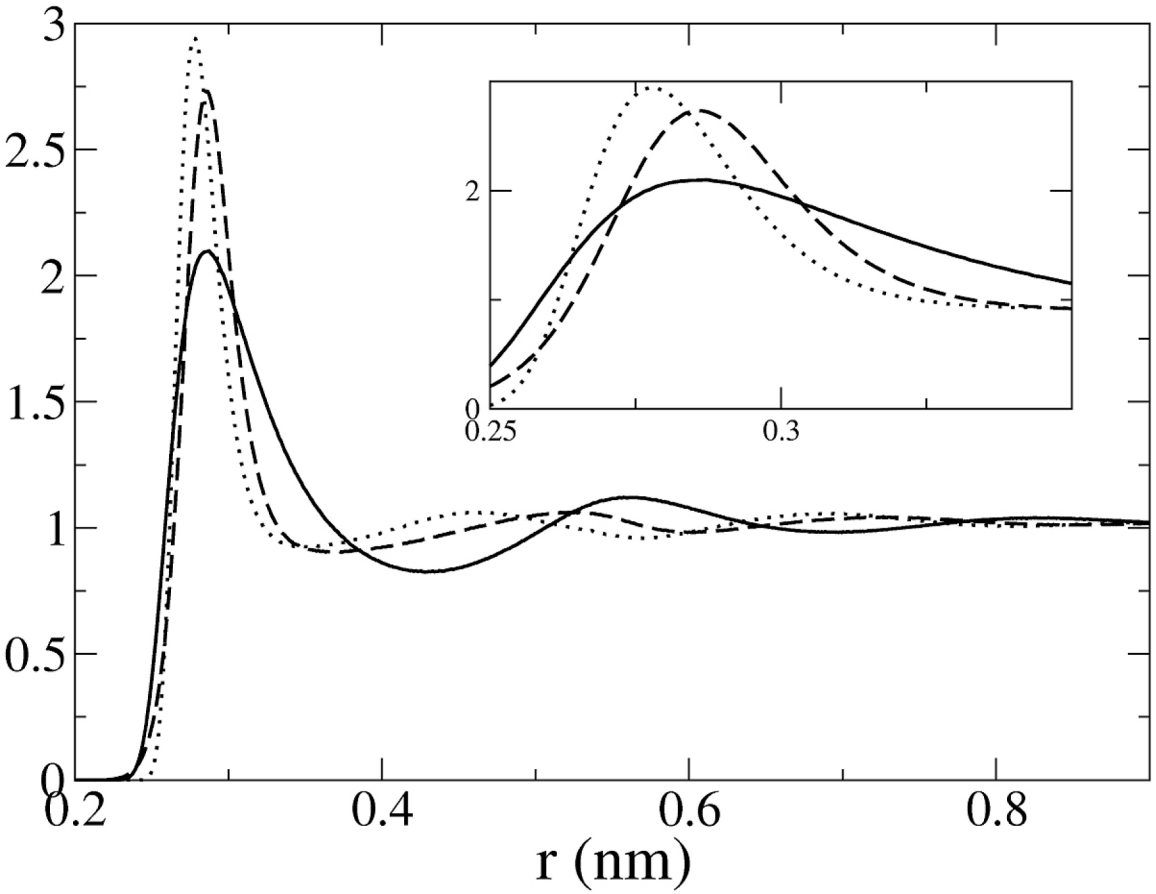


Figure 2.
The radial distribution functions (rdf's) are presented for atomistic simulations of the LJ (solid), SPC-COG (dashed), and SPC-COM (dotted) systems. The inset provides a detailed comparison of the first peak for each rdf.

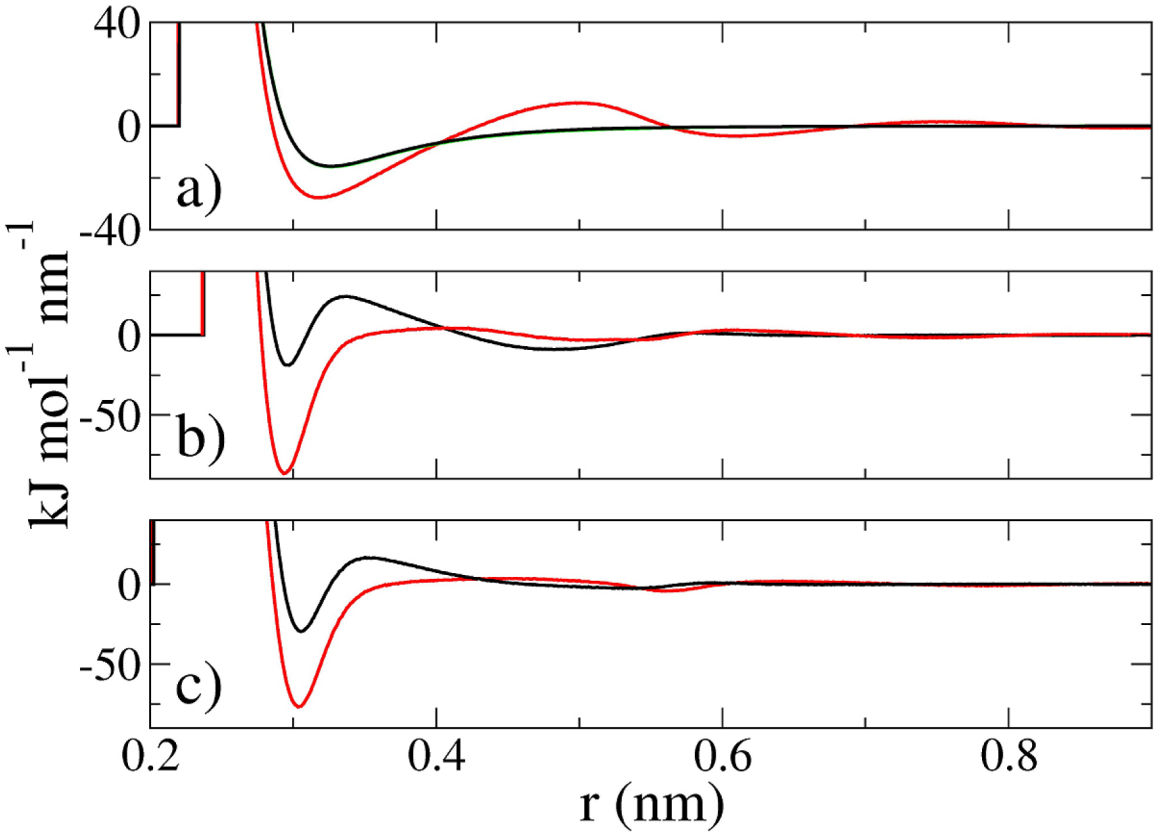


Figure 3.
The MS-CG force field (black curve) is compared with the mean force (red curve) for the LJ (panel a), SPC-COG (panel b), and SPC-COM (panel c) systems. The difference between the MS-CG force and the mean force arises from the incorporation of three-particle correlations in the MS-CG method according to eq (16). The MS-CG force field for the LJ system is also compared with the exact LJ force field (light green curve) in panel (a) and the difference is seen to be within the thickness of the curve.

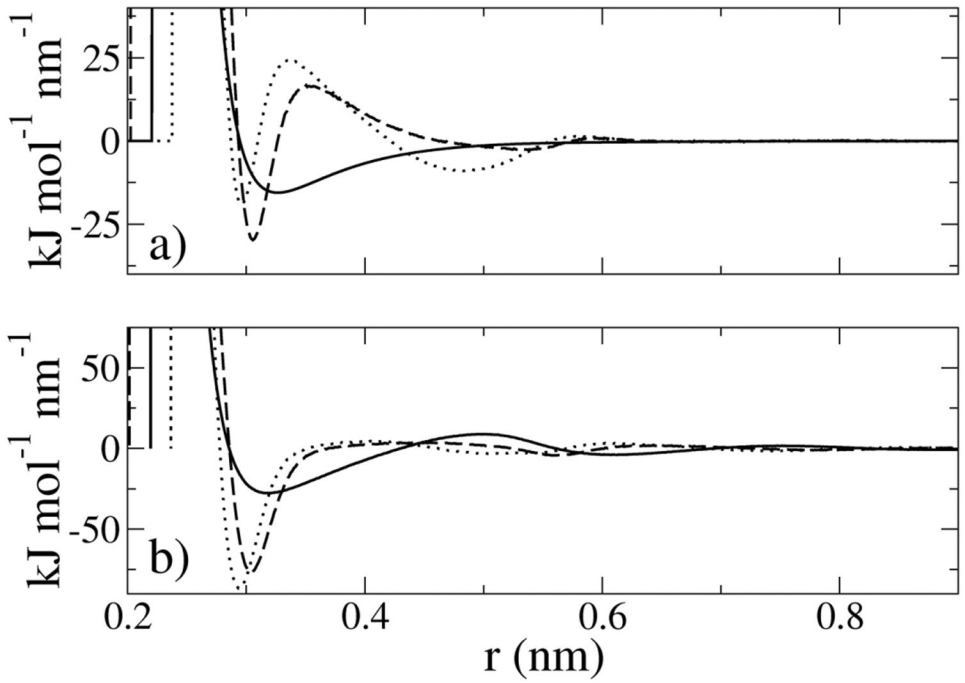


Figure 4.
The MS-CG force fields (panel a) and the mean force fields (panel b) are compared for the three different systems. In each panel the LJ results are presented as the solid curve, while the SPC-COG (COM) are presented as the dashed (dotted) curves.

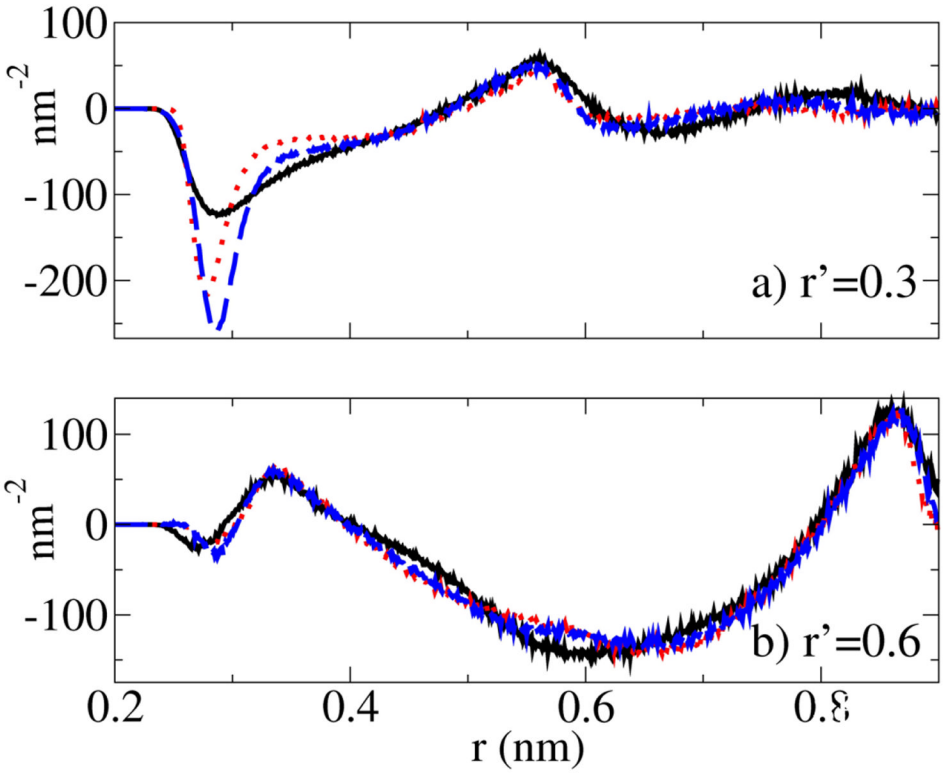


Figure 5.

Three-particle correlations between CG sites enter into the MS-CG eq (12) through the quantity $G^{(3)}(r,r')$, which is presented as a function of r for fixed $r' = 0.3$ nm (panel a) and $r' = 0.6$ nm (panel b). Each panel presents this quantity for the LJ (solid black), SPC-COG (dashed blue), and SPC-COM (dotted red) systems.