Coarse-grained force field for simulating polymer-tethered silsesquioxane self-assembly in solution

Elaine R. Chan[a)]
Department of Chemical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2136

Alberto Striolo,[b)] Clare McCabe,[c)] and Peter T. Cummings[d)]
Department of Chemical Engineering, Vanderbilt University, Nashville, Tennessee 37235-1604

Sharon C. Glotzer[e)]
Department of Chemical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2136 and Department of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan 48109-2136

(Received 2 March 2007; accepted 6 June 2007; published online 18 September 2007)

A coarse-grained model has been developed for simulating the self-assembly of nonyl-tethered polyhedral oligomeric silsesquioxane (POSS) nanoparticles in solution. A mapping scheme for groups of atoms in the atomistic molecule onto beads in the coarse-grained model was established. The coarse-grained force field consists of solvent-mediated effective interaction potentials that were derived via a structural-based coarse-graining numerical iteration scheme. The force field was obtained from initial guesses that were refined through two different iteration algorithms. The coarse-graining scheme was validated by comparing the aggregation of POSS molecules observed in simulations of the coarse-grained model to that observed in all-atom simulations containing explicit solvent. At 300 K the effective coarse-grained potentials obtained from different initial guesses are comparable to each other. At 400 K the differences between the force fields obtained from different initial guesses, although small, are noticeable. The use of a different iteration algorithm employing identical initial guesses resulted in the same overall effective potentials for bare cube corner bead sites. In both the coarse-grained and all-atom simulations, small aggregates of POSS molecules were observed with similar local packings of the silsesquioxane cages and tether conformations. The coarse-grained model afforded a savings in computing time of roughly two orders of magnitude. Further comparisons were made between the coarse-grained monotethered POSS model developed here and a minimal model developed in earlier work. The results suggest that the interactions between POSS cages are long ranged and are captured by the coarse-grained model developed here. The minimal model is suitable for capturing the local intermolecular packing of POSS cubes at short separation distances. © *2007 American Institute of Physics*.
[DOI: 10.1063/1.2753493]

I. INTRODUCTION

Self-assembly is a promising route for tailoring the organization of nanoparticles into ordered structures in a precise manner. These structures are valuable precursors for new and enhanced nanoscale materials and devices with unique properties. However, fabrication of these materials and devices requires knowledge of the atomic- and nanoscale processes that occur during the self-assembly process. In conjunction with experiments, theory and simulation are useful tools for probing self-assembly in nanoscale systems, because they afford efficient and systematic exploration of the vast parameter space and allow access to the pertinent length and time scales. The development and application of multiscale modeling and simulation techniques that bridge the various length and time scales involved in nanoscale self-assembly are receiving considerable interest lately.

Polyhedral oligomeric silsesquioxane (POSS) molecules are merely one example of the many nanoparticles presently available. The hybrid organic/inorganic character of these molecules renders them attractive for constructing nanostructured materials with useful properties. Several experiments have demonstrated that polymer-tethered POSS molecules and POSS/polymer pendant copolymers self-assemble into lamellar, cylindrical, and micellar structures in the melt or solution states.[1] The types of bulk structures that can arise from self-assembly of polymer-tethered POSS in solution and the influence of various parameters such as concentration and temperature on the formation of specific structures have been examined recently via molecular simulation.[2,3] In those simulations a minimal model of tethered POSS was devel-

[a)]Present address: Semiconductor Electronics Division, Electronics and Electrical Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899-8120. Electronic mail: elaine.chan@nist.gov
[b)]Present address: School of Chemical, Biological and Materials Engineering, The University of Oklahoma, Norman, OK 73019-1004. Electronic mail: astriolo@ou.edu
[c)]Electronic mail: c.mccabe@vanderbilt.edu
[d)]Electronic mail: peter.cummings@vanderbilt.edu
[e)]Electronic mail: sglotzer@umich.edu

oped and employed to study self-assembly of hundreds and thousands of the molecules at the mesoscale. Simulation of the corresponding numbers of POSS molecules in solvent at the explicit atom level is computationally prohibitive because thousands of POSS molecules, in a hexane solvent for example, can amount to hundreds of thousands of atoms. Furthermore, the inclusion of atomistic detail results in reduced simulation times compared to the mesoscale simulations, and hence self-assembled structures that may form on longer time scales may not be observed. Appropriate force fields also need to be developed to describe hybrid organic/inorganic molecules such as tethered POSS monomers. Despite these limitations, it has been demonstrated that standard force fields are sufficiently accurate to describe systems containing POSS monomers,[4] and detailed all-atom (AA) molecular simulation results have been reported for POSS monomers dissolved in some common organic solvents.[5–7] While those simulations are helpful for understanding effective POSS-POSS pair interactions in solution, and how these effective interactions change as temperature and solution composition vary, it remains generally unclear how to relate the parameters employed in minimal models of POSS systems to the properties obtained from AA molecular dynamics simulations.

To bridge the gap in length and time scales spanned in coarse-grained (CG) and AA simulations, it is necessary to develop mapping schemes that relate the CG models to AA results. The CG models can then be used to study self-assembly of bulk structures at long length and time scales. Presented herein is the development of a CG force field for accurately simulating the self-assembly of monotethered POSS molecules in an organic solvent. The force field consists of effective solvent-mediated interaction potentials that already account for POSS-solvent molecule interactions so that the solvent molecules do not need to be accounted for explicitly in the CG simulations. Our effort builds upon recent results obtained for linear molecules such as polymers in solution, but we extend those methods here to cubic molecules such as POSS monomers.

Current coarse-graining approaches strive to improve the computational efficiency of a simulation by systematically reducing the number of degrees of freedom in the system.[8–10] Usually, CG models provide between two and four orders of magnitude speed up in CPU time compared to AA models for similar systems. The currently available CG methodologies usually involve two steps: (1) physically mapping a detailed molecular representation onto a CG one and (2) deriving the appropriate CG interaction potentials.

With regard to the mapping procedure, two schemes are commonly employed. The first scheme involves mapping specific groups of atoms, such as one repeat unit of a polymer chain, onto CG particles. The second scheme involves mapping whole molecules, such as polymer chains, onto single CG particles. The choice between the two procedures usually depends on the level of molecular detail (e.g., connectivity) that should be preserved in the CG representation, the physical properties (e.g., intermolecular packing) that should be captured in the mesoscale simulations, and the amount of computational savings that should be afforded by the mapping. We focus here on the first method.

With regard to the development of CG effective potentials, the potentials must accurately reproduce at the mesoscale certain phenomena observed in AA simulations. Two types of potentials are often employed at this stage of the CG procedure, namely, analytical potentials with tunable parameters or numerical potentials in tabulated form. The analytical potentials typically consist of variations of phenomenological potentials, such as those that have been widely used to successfully model polymeric systems.[9,11] These potentials are parameterized according to experimental data or quantum mechanical calculations, and optimization schemes have been devised to obtain appropriate parameter values.[12] Unfortunately, the processes available to obtain the correct parameter values can be time consuming, and hence most current CG models utilize numerical potentials to describe complex interactions.

The numerical potentials can be developed by imposing that the mesoscale simulations reproduce specific intra- and intermolecular probability distribution functions obtained from the underlying AA simulations.[13–15] These structural-based coarse-graining schemes require iterative numerical methods, and are hence attractive because they can be automated. However, the resulting effective potentials are limited by their nontransferability across thermodynamic state space and inability to capture thermodynamic properties correctly, because the CG Hamiltonians are only parametrized to reproduce structural correlations appropriately.[16] Recently, an alternative numerical "force-matching" method that allows for the determination of effective force fields from any given trajectory or force data set obtained from different types of molecular dynamics simulations, including *ab initio*, path integral, and classical atomistic simulations has been developed.[17] The CG forces are cast as spline interpolations of the underlying atomistic forces and least-square fitting procedures are performed to obtain the parameters of the spline functions. This method does not require postaveraging of entire simulation trajectories and thus can be used on the fly to capture phenomena such as chemical reactions in CG molecular dynamics simulations.[18] However, this approach is limited by the difficulty associated with determining the optimal spline functions. Another class of coarse-graining techniques that involves mesoscopic blob modeling[19] has been widely used to simulate soft matter systems such as polymers and colloid-polymer mixtures.

The coarse-graining methodologies briefly summarized above have not been applied yet to systems of molecular nanoparticles with exotic geometries or tethered nanoparticles. Although CG models of nanoparticles as soft blobs are attractive from a computational viewpoint, the current methods for modeling polymer chains in this fashion have some limitations. The coarse-graining approach undertaken in this study is a structural-based one that involves deriving effective numerical potentials that will reproduce in the CG simulation target structural features in the underlying AA simulations.

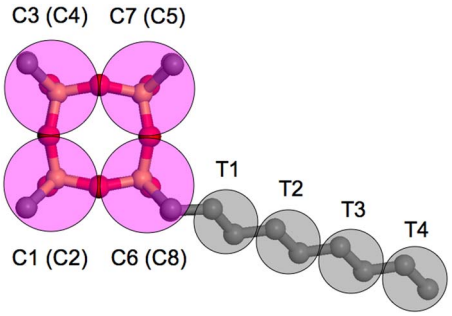

FIG. 1. (Color online) Mapping of the CG POSS molecule onto its atomistic counterpart. The silicon, oxygen, and carbon atoms are denoted in orange, red, and gray, respectively. The hydrogen atoms have been omitted for clarity. CG bead labels in parentheses correspond to beads in the background (not shown). The tether is attached to bead C8.

II. COARSE-GRAINING METHODOLOGY

A. Physical mapping of the CG model

We focus on a POSS molecule functionalized with a single nonyl tether and nonreactive methyl groups on the remaining seven silicon atoms (Fig. 1). The hydrocarbon functionalities render the molecule soluble in chemically similar and common solvents such as hexane. Because the silsesquioxane core is symmetric, one possible approach is to model the cage as a rigid cube with interaction sites on the corners as in our previous minimal model.[2] Each of the resulting eight cube corner beads encompasses one silicon atom, the neighboring oxygen atoms, and the methyl (or methylene in the case of the nonyl tether) substituent attached to the silicon atom. The beads are connected by rigid bonds and the bead interaction sites occur along the centers of the silicon-carbon bonds connecting each substituent to the cage.

To verify the validity of this CG model of the silsesquioxane cage, the bond lengths and bond angles are compared with those computed in AA simulations of nonyl-tethered POSS molecules dissolved in hexane.[7] The bond angles between CG cube corner bead sites computed in the atomistic simulations are shown in Fig. 2(a). The distribution is sharply peaked about 90°, thereby indicating that the grouping of atoms on the silsesquioxane cage is commensurate with a rigid cube structure. The distances l between the centers of the silicon-carbon bonds corresponding to the interaction sites of the cube corner beads in the CG model are also calculated. Figures 2(b) and 2(c) show that the distributions exhibit peaks centered at about l=4.2 Å. The AA molecular dynamics results indicate that it is realistic to assume that the POSS monomers are rigid cubes with eight corner sites. Mapping the AA simulation results to the CG model permits us to establish a length scale in the CG simulations by specifying the edge of the CG cube equal to this value. Each cube corner bead in the model is thus assigned a diameter of σ_c=4.2 Å.

To model the nonyl tether, two methylene groups are assigned to each CG tether bead. This mapping is on a finer scale compared to previous CG models of hydrocarbon chains that have employed groupings of three or more methylene groups per CG bead,[16,20] but such a mapping could facilitate future efforts to fully bridge length and time scales

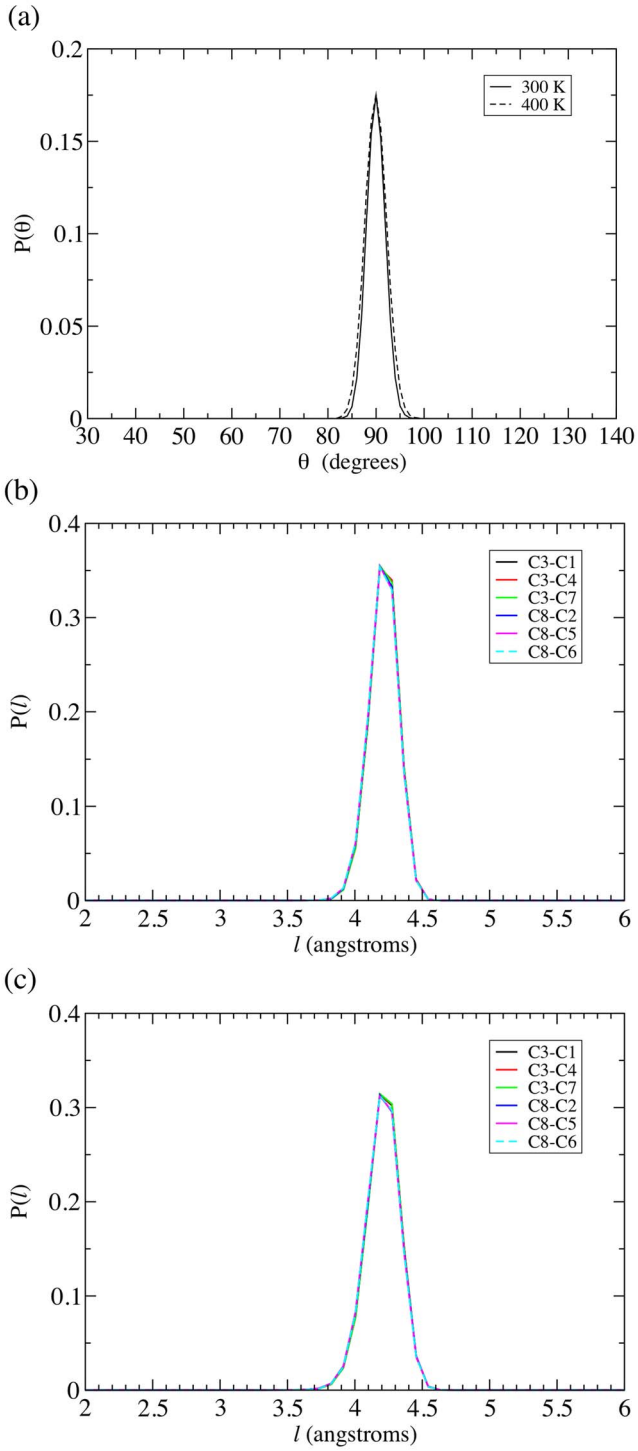

FIG. 2. (Color online) (a) C3-C7-C5 bond angle probability distribution computed from AA simulations. [(b) and (c)] Bond length probability distributions between CG cube bead sites from AA simulations at T=300 K and T=400 K, respectively.

in polymer-tethered POSS self-assembly by reverse mapping the CG model back onto its explicit atom counterpart. Note that the end tether bead actually represents a CH_2–CH_3 group in this mapping scheme, and it is assumed that the behavior and physical properties of this group do not significantly deviate from that of a CH_2–CH_2 group. The interaction sites for these beads occur along the center of the corresponding carbon-carbon bond in the AA representation.

The bond length distributions between pairs of tether bead sites are computed from AA simulations,[7] thereby allowing a length scale for these CG beads to be established. These distributions exhibit two peaks about $l=2.3$ Å and $l=2.6$ Å (as shown in Figs. 5 and 6, which will be discussed later). The two distances most likely correspond to *trans* and *gauche* configurations for the alkane chain. We assign a diameter of $\sigma_t=2.5$ Å to each bead in the CG representation of the tether.

B. Derivation of solvent-mediated effective potentials

1. Approach

We reproduce in the mesoscale simulations a selected set of target structural quantities computed from the underlying AA simulations, namely, the intramolecular bond length and bond angle probability distributions and the intermolecular radial distribution function (RDF) between cube corner beads on different molecules, that correspond to the bead interaction sites in the CG model. These quantities are likely to be the most important factors influencing the local intermolecular packing within self-assembled structures and thus the subsequent formation of specific types of bulk structures at longer length and time scales.

The algorithm employed to derive the effective potentials for nonyl-tethered POSS in hexane solvent is a numerical scheme that yields effective potentials capable of reproducing target structural correlations via the following equation:[13,15,16]

$$U_{i+1}(x) = U_i(x) + \alpha k_B T \ln\left[\frac{P_i(x)}{P_{\text{target}}(x)}\right], \quad i = 0,1,2,\ldots, \tag{1}$$

where i denotes the iteration step number, k_B is Boltzmann's constant, T is the temperature, x is the independent variable, and $P(x)$ represents a probability distribution function such as a RDF, bond length probability distribution, or bond angle probability distribution. The algorithm involves updating trial effective potentials $U_i(x)$ by successively adding correction terms that account for the deviation between the trial probability distribution function in the CG simulations and the corresponding target distribution function from the AA simulations. The parameter α is an arbitrary number that is chosen to appropriately scale the magnitude of the correction term so that stable convergence of the effective potentials is attained. An effective CG potential that reproduces the same structural features from the corresponding AA simulations is obtained when the trial and target distribution functions are equal or nearly equal within some prescribed tolerance value.

The concept of structural-based coarse graining is motivated by the proof that for simple pairwise additive and spherically symmetric potentials, there is a unique mapping between the RDF and the intermolecular potential at a given thermodynamic state point.[21] Equation (1) is inspired by the relationship between the potential of mean force (PMF) and the RDF for molecular centers of mass. At infinite dilution the PMF is given by the expression[22]

$$U_{\text{PMF}}(r) = -k_B T \ln[g(r)] \tag{2}$$

and is exactly equal to the intermolecular pair potential between two point particles. It is important to note here that Eq. (2) for the PMF is strictly applicable to particles or molecules with single interaction sites, and that there is a common misconception that this expression is valid between multiple interaction sites on molecules, such as polymer chains, in which orientation correlations must be accounted for (see Appendix for further details). Because the CG tethered POSS model in this study contains multiple interactions sites (i.e., beads), Eq. (1) is utilized here with the understanding that it is simply one of many possible numerical algorithms capable of yielding mesoscale effective potentials by satisfying the boundary condition that the trial mesoscale potentials must converge when the corresponding trial distribution function matches the target distribution function from the atomistic simulations. Equation (1) is an elegant algorithm because the logarithmic correction term is able to change sign ($\pm$) accordingly so that the updated effective potential produces a trial distribution function in closer agreement with the target distribution function.

To generate the initial guesses ($i=0$) for the effective CG potentials, the target RDF, bond length probability distribution $P(l)$, and bond angle probability distribution $P(\theta)$ from the atomistic simulations are Boltzmann inverted via the following equations, respectively:

$$U_0(r) = -k_B T \ln[g_{\text{target}}(r)], \tag{3}$$

$$U_0(l) = -k_B T \ln[P_{\text{target}}(l)], \tag{4}$$

$$U_0(\theta) = -k_B T \ln\left[\frac{P_{\text{target}}(\theta)}{\sin\theta}\right]. \tag{5}$$

The sine term in Eq. (5) arises from the Jacobian transformation between spherical and Cartesian coordinates. Note that these choices for the initial guesses are rather arbitrary, as discussed further below in Sec. II B 2 and shown in the Appendix.

To assess convergence of the derived effective potentials, the following merit functions[10,15] or error integrals are computed during each iteration step:

$$f_{\text{merit,RDF}} = \int w(r)[g_i(r) - g_{\text{target}}(r)]^2 dr, \tag{6}$$

$$f_{\text{merit,bond}} = \int w(l)[P_i(l) - P_{\text{target}}(l)]^2 dl, \tag{7}$$

$$f_{\text{merit,angle}} = \int [P_i(\alpha) - P_{\text{target}}(\alpha)]^2 d\alpha. \tag{8}$$

Equations (6)–(8) are the merit functions corresponding to the intermolecular cube corner bead RDF, intramolecular bond length probability distributions, and intramolecular bond angle probability distributions, respectively. Optional non-negative weighting functions $w(r)=\exp(-r/\sigma_c)$ and $w(l)=\exp(-l/\sigma_t)$ are included to penalize deviations between

the CG trial distribution function and the AA target distribution function at small separation distances.

On the basis of RDFs computed for the CG tether bead sites from the AA simulations,[7] a purely repulsive soft-sphere potential[23] is used to capture the intermolecular excluded volume interactions between tether beads

$$U(r) = \frac{27\varepsilon}{4}\left[\left(\frac{\sigma}{r}\right)^9 - \left(\frac{\sigma}{r}\right)^6\right] + \varepsilon, \quad r \leqslant r_c,$$

$$U(r) = 0, \quad r > r_c, \tag{9}$$

where $r_c = (3/2)^{1/3}\sigma$ and $\varepsilon = k_B T$. The choice of this potential does not significantly affect the CG probability distribution functions involving the tether beads nor self-assembly of the molecules.

2. Alternative routes

It is possible that different effective potentials exist that can each reproduce the target distribution function from the AA simulations. One way to assess the accuracy of the CG potentials is to derive them from different types of initial guesses for the potential using the same numerical algorithm and then compare the results. This approach can aid in corroborating the soundness of an effective potential if different initial guesses yield the same result or allow one to assess the best effective potential if different ones arise from the derivation process.

We evaluate the accuracy of the intermolecular bare POSS cube corner bead effective potentials by first deriving the potentials from initial guesses equal to Boltzmann inversions of the target RDFs from the AA simulations per Eq. (3). As shown in the Appendix, there is no strong theoretical basis for using this as the initial guess. Hence, we investigate what types of effective potentials arise from a different initial guess equal to the purely repulsive Weeks-Chandler-Andersen[24] (WCA) interaction potential

$$U(r) = 4\varepsilon\left[\left(\frac{\sigma_c}{r}\right)^{12} - \left(\frac{\sigma_c}{r}\right)^6\right] + \varepsilon, \quad r \leqslant r_c,$$

$$U(r) = 0, \quad r > r_c, \tag{10}$$

where $r_c = 2^{1/6}\sigma_c$ and $\varepsilon = k_B T$.

We are also interested in comparing the effective potentials generated by different numerical equations, since as discussed in the Appendix, the iterative scheme of Eq. (1) has no strong theoretical basis. As the logarithmic correction term in Eq. (1) is able to change sign ($\pm$) accordingly so that the updated effective potential produces a distribution function in the CG simulations that is in better agreement with the target distribution function from the AA simulations, this property of the correction term can serve as one criterion for formulating alternative numerical algorithms that are equally or potentially superior to Eq. (1) for deriving effective potentials. A simple correction term that satisfies both this criterion and the boundary condition that convergence of the effective potential occurs when the CG and target RDFs are equal is one that takes the linear difference between these two RDFs. Consequently, we propose the following numerical algorithm for deriving effective potentials:

$$U_{i+1}(r) = U_i(r) + \alpha k_B T[g_i(r) - g_{\text{target}}(r)], \tag{11}$$

where α has the same meaning as in Eq. (1). We compare the effective potentials generated by Eqs. (1) and (11) from identical initial guesses $U_0(r)$ and also compare the speed of each algorithm.

3. Simulation details

A stochastic molecular dynamics simulation method is used in the CG simulations. The equation of motion for each bead i is

$$m_i \dot{\mathbf{v}}_i(t) = -m_i \xi_i \mathbf{v}_i(t) + \mathbf{F}_i(\mathbf{x}_i(t)) + \mathbf{R}_i(t), \tag{12}$$

where m_i is the mass of bead i, and $\mathbf{x}_{i_i}$ $\mathbf{v}_i$, $\mathbf{F}_i$, and ξ_i represent the position, velocity, force, and friction coefficient acting on bead i, respectively.[25] These simulations sample the canonical ensemble, as the friction coefficient and stochastic noise term couple the system to a heat bath and effectively function together as a nonmomentum conserving thermostat. Systems of $Nb=5$ and $Nb=20$ CG nonyl-tethered POSS molecules ($N=40$ and 240 total particles, respectively) are simulated using cubic simulation boxes and periodic boundary conditions at overall density $\rho=0.75$ g/cm^3 and temperatures $T=300$ and 400 K. The equations of motion are integrated using the leap-frog algorithm.[24] The rigid-body motion of the cube is captured using the method of quaternions.[24] Initial configurations are generated by relaxing each system athermally (i.e., excluded volume interactions only). The structural evolution of each system over time is monitored by inspecting snapshots of configurations.

These configurations are compared to those generated from the corresponding AA molecular dynamics simulations having the same number of molecules and at the same values of temperature and density. AA simulations of $Nb=20$ nonyl-tethered POSS molecules and 987 hexane chains ($N=6642$ total atoms) at overall density $\rho=0.75$ g/cm^3 and temperatures $T=300$ and 400 K are performed using the DL_POLY (Ref. 26) simulation package. The Frischknecht-Curro force field[27] is used to describe the POSS cage and the TRAPPE force field[28] is used to describe the alkane tethers and the hexane solvent. Additional details of the atomistically detailed simulations can be found in Ref. 7.

III. RESULTS AND DISCUSSION

A. "Bare" POSS molecules

We first derive an intermolecular effective potential that captures the interactions between cube corner beads. This is a logical starting point since the addition of a single hydrocarbon tether on one corner of the silsesquioxane cage does not impact the behavior of the cage significantly.[29] Thus, it is anticipated that the tether will have a minimal impact on the intermolecular interactions between nonreactive, octafunctionalized "bare," or nontethered, POSS molecules.

AA molecular dynamics simulations of $Nb=5$ octamethyl functionalized POSS molecules dissolved in hexane have been performed at overall density $\rho=0.75$ g/cm^3 and

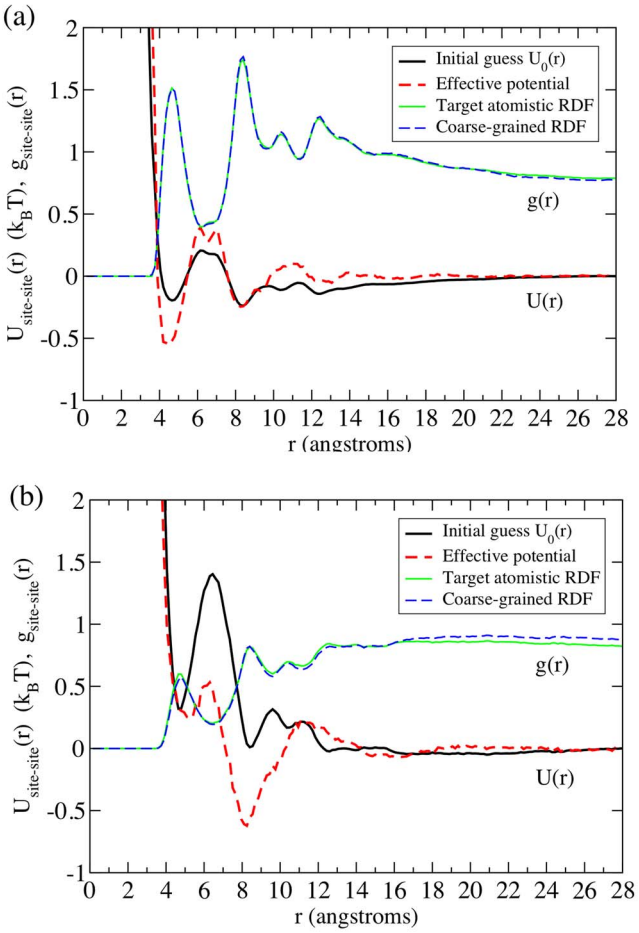

FIG. 3. (Color online) Cube corner bead-corner bead effective potentials and intermolecular radial distribution functions for bare POSS molecules. (a) $T=300$ K and (b) $T=400$ K.

temperatures $T=300$ and 400 K.[7] Target RDFs characterizing the local structure among the CG cube corner bead sites from the corresponding atomistic molecules are computed from these simulations and shown in Fig. 3. The RDFs exhibit multiple pronounced peaks that occur primarily at integer values of the cube edge length at lower temperature, as observed for POSS monomers in hexadecane.[5] Note that the tails in the RDFs at large separation distances fall below unity at both temperatures. This result is attributed to a combination of three factors: (1) effects of small system size ($Nb=5$ molecules or $N=40$ particles), which is rectified by multiplying the RDF by a correction factor $N/(N-1)$,[22,30] (2) normalization of the RDF does not account for the close proximity of cube corner bead sites arising from the rigid bond constraints, and (3) nonuniform clustering of POSS cubes throughout the simulation box.

The CG effective pair potentials derived on the basis of these RDFs are also shown in Fig. 3, along with the corresponding initial guesses used in the iteration algorithm, for both temperatures. An interaction potential cutoff value $r_c=28$ Å that is slightly less than half the simulation box length is employed. Small correction steps ($\alpha=0.01-0.1$) are necessary during the iteration process to stabilize convergence of the potentials. Previous applications of the algorithm to derive effective potentials for polymer melts report success with larger values $\alpha=0.2$ (Ref. 16) and $\alpha=1$.[15] Elimination of the solvent molecules in the CG model likely requires that smaller correction steps be taken to stabilize convergence of the solvent-mediated effective potentials.

At $T=300$ K the CG potential exhibits an attractive well at $r=4.7$ Å followed by a broad repulsive peak centered at $r=7$ Å. The potential at larger separation distances consists of an alternating series of attractive wells and repulsive peaks corresponding to the peaks and valleys, respectively, in the target RDF from the AA simulations. At higher temperature $T=400$ K the effective pair potential exhibits a steep attractive well at $r=8.3$ Å. At larger separation distances $r>8$ Å, the potential exhibits attractive wells and repulsive peaks that are attenuated to a greater extent than those observed at lower temperature. This behavior is expected as the RDF indicates loss of long-range structure with increasing temperature.

The RDFs generated by the effective pair potentials in the CG simulations are shown in Fig. 3. There is excellent agreement between the CG and target RDFs at $T=300$ K, and this result is quantitatively corroborated by merit function values on the order of $f_{\text{merit,RDF}}\sim 10^{-5}$ when convergence is attained.[15] At $T=400$ K there is good agreement between the RDFs, although the CG RDF tail at separation distances $r>16$ Å does not quantitatively match the atomistic RDF tail precisely. The corresponding merit function values are on the order of $f_{\text{merit,RDF}}\sim 10^{-4}$ when the iteration algorithm converges. To further test the accuracy of the coarse-graining procedure, we compute the RDFs between the centers of mass of the POSS molecules from the CG model and the underlying atomistic model. Figure 4 displays the cube center of mass RDFs for both temperatures considered. The CG simulations exhibit a more pronounced structure than the underlying atomistic simulations but there is good overall agreement between the shape and magnitude of the curves, especially when we consider that the cube centers of mass are not interaction sites in the CG model. The increased local structure exhibited in the CG simulations can be attributed to the rigidity of the cubes in the model.

B. Monotethered POSS molecules

We next derive effective potentials to capture bond stretching and bond bending interactions in the monotethered POSS CG model. We build upon the model developed thus far for bare cubes by considering the interactions introduced when a nonyl tether is attached to one corner (Fig. 1). Because the POSS cages are treated as rigid cubes in the model, only four bonded pairs of beads must be considered—one bond between the cube corner bead and the adjacent tether bead (C8-T1) and three bonds between adjacent pairs of tether beads (T1-T2, T2-T3, and T3-T4).

The AA simulated probability distributions used to derive the effective bond stretching potentials are shown in Figs. 5 and 6. Further details of these simulations are reported elsewhere.[7] The most probable bond length values are essentially independent of the temperatures studied here. These results are sensible because each tether bead represents only two methylene groups; hence, the bond length

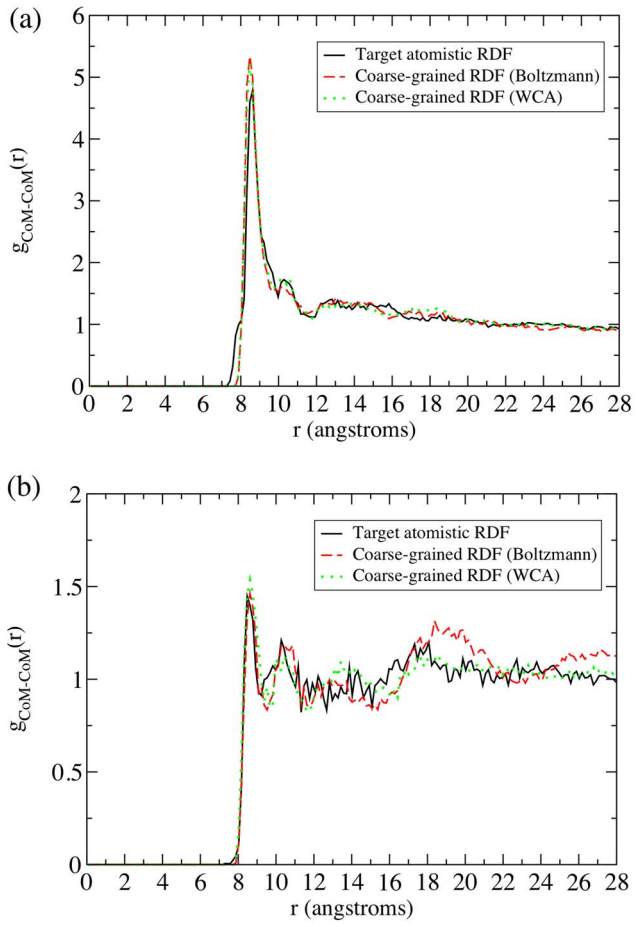

FIG. 4. (Color online) Intermolecular radial distribution functions between center of mass positions for bare POSS molecules. CG results obtained from different initial guesses for the potentials are displayed. (a) $T=300$ K and (b) $T=400$ K.

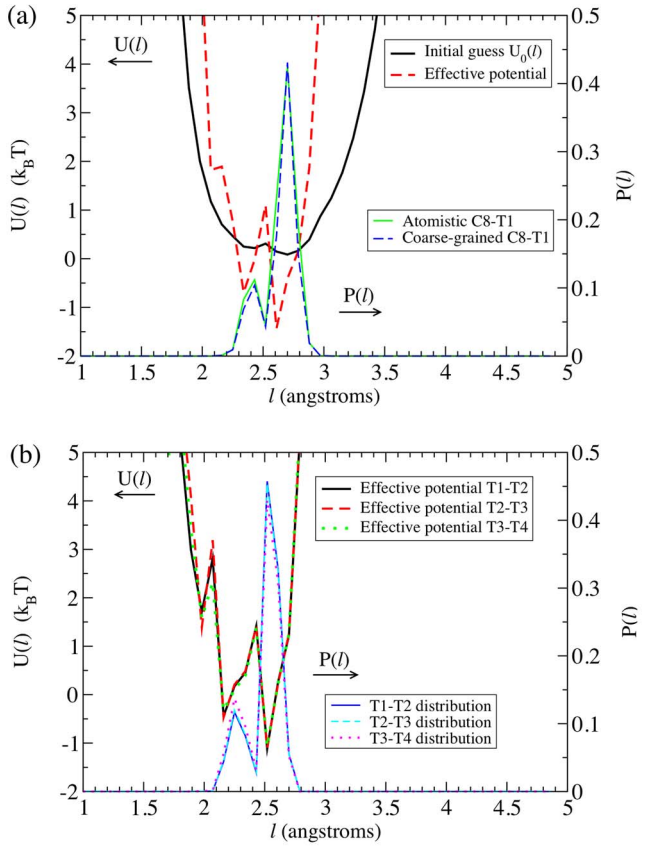

FIG. 5. (Color online) Effective bond stretching potentials and the corresponding probability distributions for CG monotethered POSS molecules at $T=300$ K. (a) Bonded cube and tether beads. (b) Bonded tether beads.

values between CG interaction sites are likely to be dictated by the chemical connectivity in the underlying atomistic model and thus should not depend strongly on temperature. The temperature influences the probabilities that specific bond length values occur during the simulation.

The converged CG effective potentials derived on the basis of these distributions are also displayed in Figs. 5 and 6 for both temperatures. A value of the parameter $\alpha=0.01$ is used to obtain stable convergence of the potentials during the iteration process. The effective bond potential for the C8-T1 bead pair exhibits two narrow wells that correspond closely to the peak positions in the target probability distributions. The potential energy is negative at the bottom of each well and indicates that the cube corner and adjacent tether beads are attracted to each other. The attraction results in bead overlap because this mechanism allows for the most probable bond lengths to be sampled given the bead sizes $\sigma_c=4.2$ Å and $\sigma_t=2.5$ Å. Bead overlap is physically allowable in the CG model since each bead encompasses groups of atoms, and this effective grouping of atoms gives rise to the two peaks observed in the target probability distribution. At each temperature the effective bond potentials for the T1-T2, T2-T3, and T3-T4 tether bead pairs are nearly identical. Each potential exhibits a series of narrow wells that mirror the shape of the target probability distributions. The CG and target distributions for the C8-T1 bead pair are in excellent agreement with $f_{\text{merit,bond}} \sim 10^{-5}$ when convergence is attained. The CG and target distributions for the T1-T2, T2-T3, and T3-T4 tether bead pairs are in good agreement with $f_{\text{merit,bond}} \sim 10^{-4}$ when convergence is attained.

Lastly, we derive the effective potentials to capture the target intermolecular bond angle probability distributions from the AA simulations in the CG monotethered POSS model. Four angles are considered here as defined by the following triplets of beads: C6-C8-T1, C8-T1-T2, T1-T2-T3, and T2-T3-T4 (Fig. 1). We do not consider dihedral interactions in the model to maintain its simplicity.

The target probability distributions corresponding to the CG bead interaction sites are shown in Figs. 7 and 8. The most probable bond angle values are fairly independent of the temperatures studied here. The multiple peaks in the distribution functions can probably be attributed to dihedral transitions along the hydrocarbon chain that are captured in the fine level of coarse-graining employed. The CG effective potentials are also illustrated in Figs. 7 and 8 for both temperatures. A value of the parameter $\alpha=0.01$ is used to obtain stable convergence of the potentials during the iteration process. At both temperatures the effective potentials consist of a series of peaks and valleys that essentially mirror the shape and relative magnitude of these features in the target probability distribution functions. The CG and target distributions for the C6-C8-T1 bond angle are in excellent agreement at

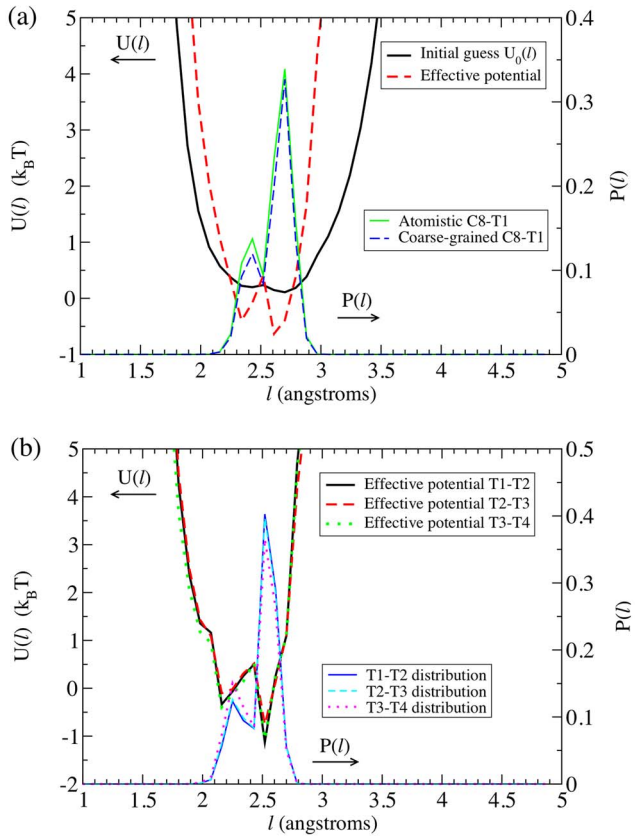

FIG. 6. (Color online) Effective bond stretching potentials and the corresponding probability distributions for CG monotethered POSS molecules at $T=400$ K. (a) Bonded cube and tether beads. (b) Bonded tether beads.

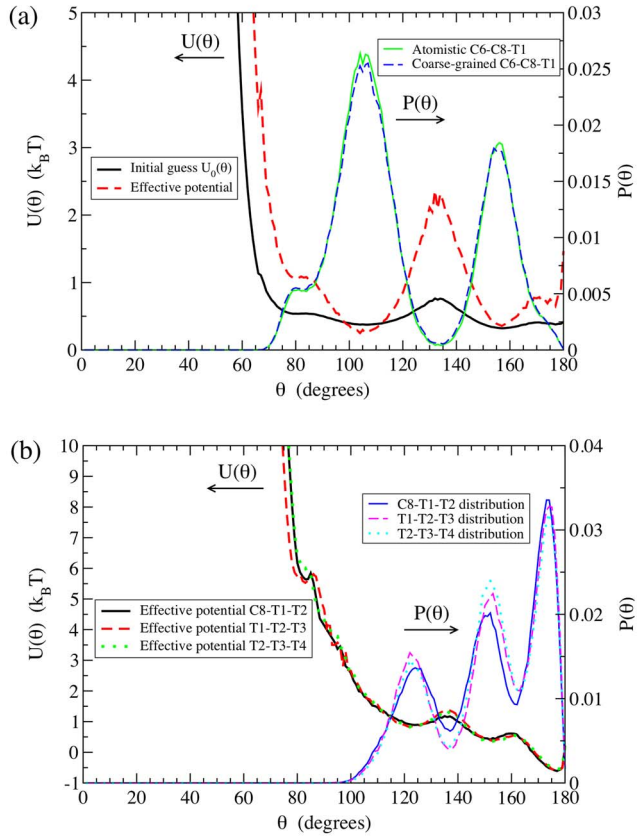

FIG. 7. (Color online) Effective bond bending potentials and the corresponding probability distributions for CG monotethered POSS molecules at $T=300$ K. (a) C6-C8-T1 angle. (b) C8-T1-T2, T1-T2-T3, and T2-T3-T4 angles.

both temperatures with $f_{\text{merit,angle}} < 10^{-6}$ when convergence is attained. The CG and target distributions for the C8-T1-T2, T1-T2-T3, and T2-T3-T4 bond angles are in good agreement at both temperatures with $f_{\text{merit,angle}} \sim 10^{-6}$ when convergence is attained.

IV. CG FORCE FIELD EVALUATION AND VALIDATION

A. Varying initial guesses for the effective potentials

As mentioned previously, the CG effective potentials derived via the structural-based coarse-graining algorithm adopted in this study are nonunique. We investigate how a different initial guess affects the resulting CG potentials for bare POSS cubes. Using the WCA potential as the initial guess, CG bare cube effective potentials are derived via Eq. (1) with $\alpha=0.01$ and displayed in Fig. 9. At $T=300$ K the effective potential appears to be essentially identical to that derived from the initial guess generated by Eq. (3). However, at higher temperature $T=400$ K the form of the effective potential obtained from the WCA initial guess differs significantly from that derived from the initial guess in Eq. (3). Figure 9 shows that the cube corner bead RDFs generated from the initial guess of the WCA potential are in good agreement with the target RDFs. The agreement between these RDFs, as well as the cube center of mass RDFs (Fig. 4), is better at $T=400$ K over the entire range of interbead separation distances when the initial WCA guess for the effective potential is used. Furthermore, the merit function values here are on the order of $f_{\text{merit,RDF}} \sim 10^{-5}$ at this temperature, and thereby quantitatively indicate better agreement between the RDFs. The WCA initial guess may result in a more accurate effective potential at $T=400$ K because the WCA potential does not have the multiple peaks and valleys that are present in the initial guess of Eq. (3). These peaks and valleys tend to make the evolution towards the desired effective potential cumbersome.

B. Varying numerical iteration algorithms

Another approach to evaluate the accuracy of the resulting CG potentials is to use the same initial guess for the effective potential but vary the form of the iterative equation used to derive the effective CG potential. This method can provide an additional measure of accuracy if two different iteration algorithms can produce identical potentials. We examine the intermolecular cube corner bead effective potentials for bare POSS cubes. The effective potentials generated by Eq. (11) with an initial guess of the WCA repulsive potential and $\alpha=0.01$ are investigated at temperatures $T=300$ and 400 K. The effective potentials obtained from numerical iteration via Eqs. (1) and (11) are plotted together in Fig. 10. The overall shapes of the potentials are essentially identical, but the linear algorithm tends to generate attractive wells and repulsive peaks of slightly different magnitudes at interbead separation distances $r<9$ Å. These differences in magnitude

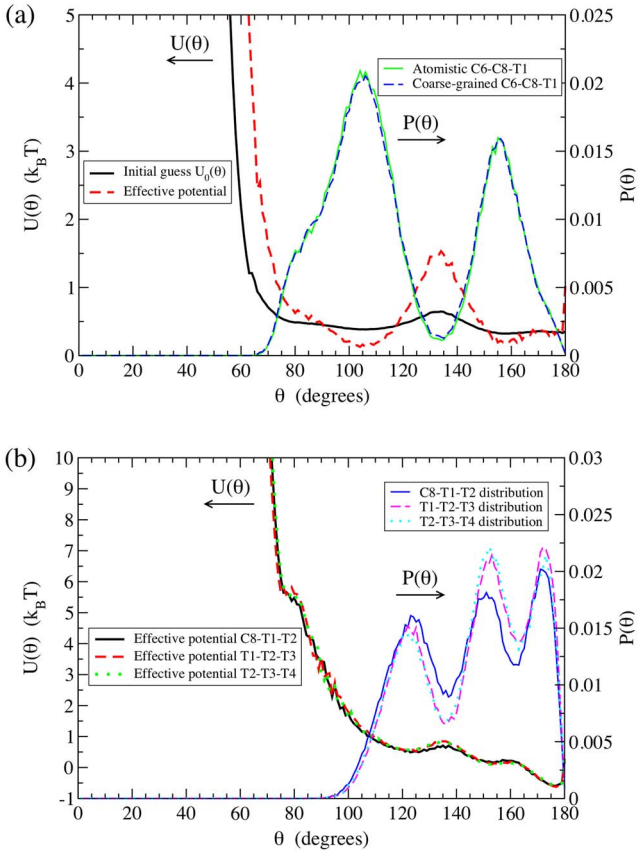

FIG. 8. (Color online) Effective bond bending potentials and the corresponding probability distributions for CG monotethered POSS molecules at $T=400$ K. (a) C6-C8-T1 angle. (b) C8-T1-T2, T1-T2-T3, and T2-T3-T4 angles.

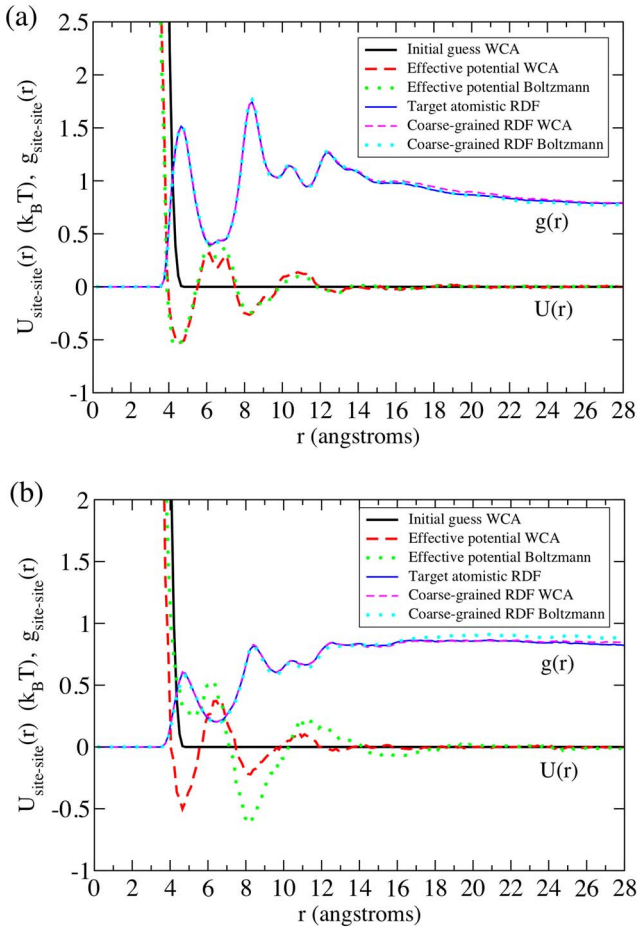

FIG. 9. (Color online) Site-site CG effective potentials for bare POSS molecules derived from different initial guesses for the potentials and the corresponding intermolecular radial distribution functions. (a) $T=300$ K and (b) $T=400$ K.

may result from how the two algorithms update the potentials at interbead separation distances where either the atomistic or CG RDF is zero. The logarithmic correction term in Eq. (1) requires that these RDFs be nonzero over the entire range of separation distances while the linear correction term in Eq. (11) exists for all RDF values. To avoid infinite values of the logarithmic correction term in Eq. (1) for zero RDF values at small separation distances, the effective potential was extrapolated. Figure 10 indicates that, while the extrapolated potential matches the shape of the corresponding portion derived via Eq. (11), the magnitudes of these parts of the potential differ slightly at identical separation distances. However, overall these results imply that the effective potentials are fairly independent of the iteration scheme employed here to derive them. The CG RDFs generated from the converged effective potentials produced from Eq. (11) are displayed in Fig. 10. There is good agreement between these RDFs and the corresponding target RDFs from the AA simulations.

We examine the number of iteration steps required for the effective potential to converge to compare the speed of Eqs. (1) and (11). Our findings indicate that convergence of the potential occurs slightly faster when Eq. (11) is utilized. At $T=300$ K convergence is achieved after 232 iterations using Eq. (11) versus 243 iterations using Eq. (1). At $T=400$ K convergence is achieved after 218 iterations via Eq. (11) while 229 iterations are required to attain convergence via Eq. (1). Each iteration step consumes 17–18 min CPU time on a single 2 GHz Apple G5 processor.

C. Validation from atomistic simulations

To examine the accuracy of the solvent-mediated CG force field for simulating monotethered POSS self-assembly, CG simulations of small systems of these molecules ($Nb=20$) are performed, and the self-assembled structures and local intermolecular packing of molecules are compared with those observed in the corresponding explicit atom simulations having the same number of molecules, temperature, and density.

In systems of bare POSS monomers small aggregates containing two to three POSS molecules are observed in both the AA and CG simulations, with aggregation being more prevalent at lower temperature. The POSS molecules within each aggregate in the AA simulations sample several relative orientations, namely, face-face, corner-corner, corner-edge, and edge-edge packings. These types of local packing of the cages were also previously observed in AA simulations of octamethyl functionalized POSS dissolved in hexadecane solution.[5] Snapshots from the CG simulations reveal that the cubes also exhibit the same types of local packing (Fig. 11).

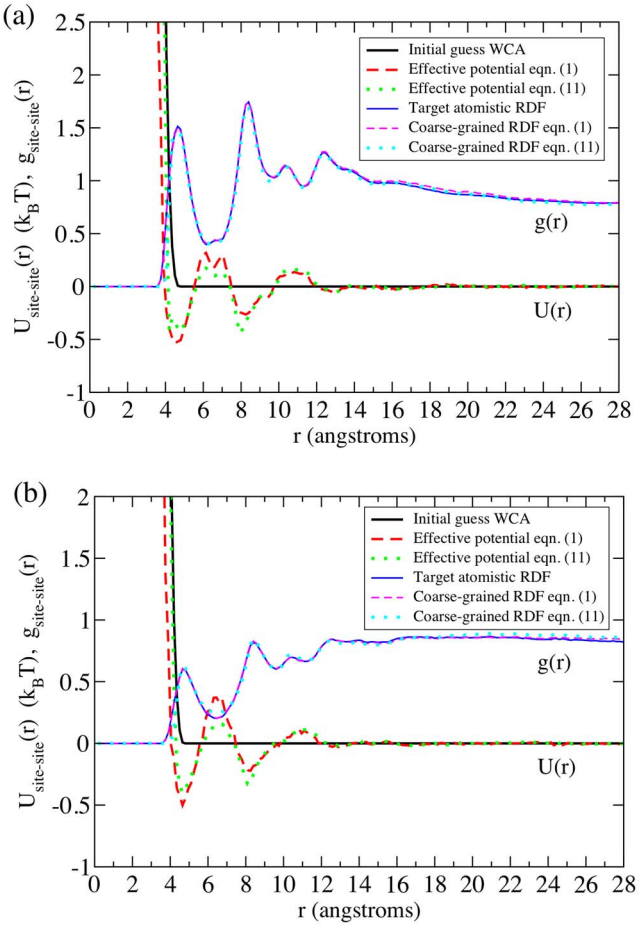

FIG. 10. (Color online) Site-site CG effective potentials for bare POSS molecules derived from different numerical iteration equations and the corresponding intermolecular radial distribution functions. (a) $T=300$ K and (b) $T=400$ K.

In both the atomistic and CG simulations, the face-face packing of cubes is the most predominant. This type of packing corresponds to the large peaks in the RDFs that occur at approximately integer values of the CG cube edge length. Previous self-assembly simulations of monotethered POSS molecules employing a minimal model also exhibited solely the face-face local packing of cubes,[3] and thus the variety of orientations captured in the CG model here indicates an improvement in model accuracy.

In systems of monotethered POSS monomers local aggregation of two to three POSS molecules is observed in both the atomistic and CG simulations. The frequency at which clusters of molecules form and subsequently dissolve is observed to be faster at higher temperature. Figure 12 contains simulation snapshots of selected aggregates. The local intermolecular packings of cubes within the aggregates are similar to those observed in the bare POSS cube simulations. The hydrocarbon tethers in both the atomistic and CG systems point outwards away from the aggregate cores of the cages. This behavior is expected as the hydrocarbon chains prefer to interact with the chemically identical solvent molecules. In both the atomistic and CG simulations, the hydrocarbon tethers appear to favor nearly perfectly linear conformations at low temperature, while at higher temperature the tethers tend to sample a broader range of bond angles, and various bent conformations are observed.

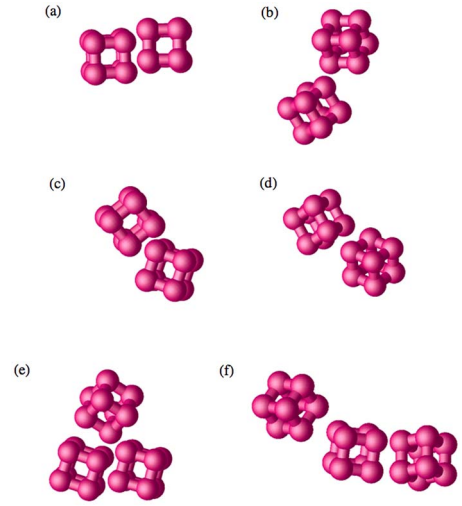

FIG. 11. (Color online) Local packings of bare CG POSS molecules: (a) face-face, (b) corner-corner, (c) corner-edge, and (d) edge-edge. Aggregates of three POSS molecules exhibit combinations of these intermolecular orientations, such as (e) corner-corner and face-face, and (f) corner-edge and edge-edge.

The behavior of the tethers in the CG simulations is quantitatively evaluated by computing the probability distribution functions for the intramolecular nonbonded distances between the end tether bead (T4) and cube corner bead to

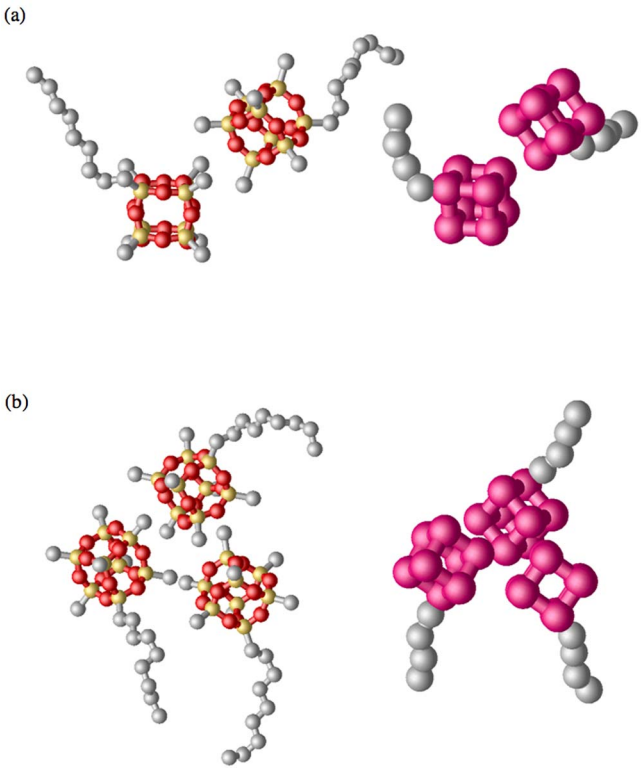

FIG. 12. (Color online) Snapshots of aggregates observed in both AA and CG simulations of monotethered POSS molecules in hexane solution. Solvent molecules have been omitted from the AA snapshots for clarity. (a) Selected atomistic and CG aggregates containing two POSS molecules. (b) Selected atomistic and CG aggregates containing three POSS molecules.

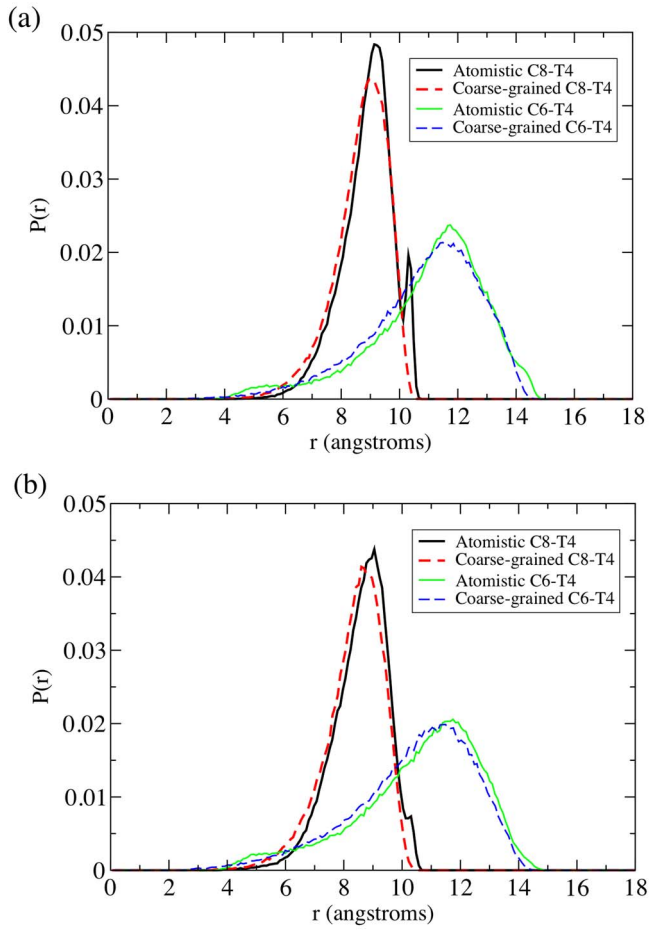

FIG. 13. (Color online) Probability distribution functions for intramolecular nonbonded distances between end tether beads and cube corner beads. (a) $T=300$ K and (b) $T=400$ K.

which the tether is attached (C8) and between the end tether bead and a nearest neighbor cube corner bead (C6) (Fig. 1). Note that effective potentials for the C8-T4 and C6-T4 interactions were not derived here to reproduce the corresponding AA target distribution functions in the CG simulations, because the effective bond stretching and bond bending potentials derived in Sec. III should, in principle, be sufficient to capture these intramolecular nonbonded distance distributions. Figure 13 shows that the shapes of the target atomistic and the CG distributions appear to be in good agreement. The relative magnitudes of the major C8-T4 and C6-T4 peak heights appear to be suitably captured in the CG simulations.

The CG model developed here affords a savings in computing time of roughly two orders of magnitude compared to the underlying atomistic simulations. The latter required nearly 420 CPU hours on a single 2 GHz Apple G5 processor to simulate 5 ns, while the corresponding CG simulations required approximately 3.5 CPU hours. The speedup in computing time for the CG model can be largely attributed to the reduced number of particles that arises because the solvent molecules are treated implicitly.

V. COMPARISONS BETWEEN CG AND MINIMAL MODELS OF MONOTETHERED POSS MOLECULES

It is instructive to compare features of the CG model of solvated monotethered POSS molecules developed here with

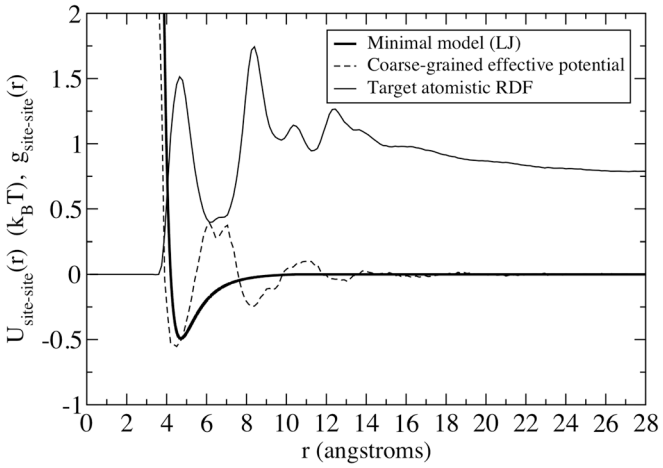

FIG. 14. Effective potentials for iteration between cube corner beads in minimal and CG models of monotethered POSS at $T=300$ K.

those of the minimal model developed previously.[2] As mentioned in the previous section, the CG model successfully captures the various types of silsesquioxane cage intermolecular orientations exhibited by the AA simulations, while the minimal model only captures face-face local intermolecular packing. The location and shape of the CG octamethyl functionalized bare POSS cube effective potential and the phenomenological Lennard-Jones (LJ) potential used to portray these interactions in the minimal model are similar with respect to the first attractive well at separation distances $r<5$ Å (Fig. 14). This short-ranged attraction appears to be primarily responsible for generating the face-face local intermolecular packing between POSS cubes in the atomistic and CG simulations, so perhaps it is understandable why only this particular type of intermolecular orientation is observed in the minimal model studies. The shape of the CG effective potential at larger separation distances past the first attractive well differs significantly from the LJ potential. The series of attractive wells and repulsive peaks at $r>10$ Å appear to be responsible for generating the corner-edge, corner-corner, and edge-edge intermolecular orientations, and this finding indicates that the interactions between bare POSS cubes are long ranged. The interaction cutoff employed in the corresponding minimal model is $r_c=2.5\sigma_c=10.5$ Å, and this value is smaller than the third and fourth peak positions in the cube corner bead RDFs that represent the other types of intermolecular packings observed in both the atomistic and CG simulations. Hence, the minimal model appears to be accurate for capturing the local intermolecular packing of POSS cubes at short separation distances $r<10$ Å. This model will most likely be unsuccessful in capturing the POSS cube intermolecular packings at larger separation distances, even if the value of the LJ potential interaction cutoff is increased accordingly, because the tail of this potential has small values $\sim 10^{-3}k_BT$ and lacks the series of attractive wells and repulsive peaks that appear to promote the additional types of intermolecular orientations.

With regard to the tether, the CG model captures the predominately extended chain conformations observed in the AA simulations as a result of the derivation and implementation of the effective bond angle potentials. These potentials

are required to faithfully capture the interactions between the tether and the POSS cage. Simulations of the CG model molecules with only the effective bare POSS cube and effective bond stretching potentials implemented do not reproduce well the orientation of the tether relative to the attached POSS cage nor the bond angles along the tether. Snapshots of these CG configurations reveal that the fully flexible tether tends to be in closer proximity to the attached cube by curving around it. Furthermore, the flexible tether adopts bent conformations that are not observed in the atomistic simulations. In the minimal model, the tether is treated as a fully flexible chain and also suffers from the same shortcomings as the CG model without specific bond bending interactions accounted for.

The neutral solvent condition appears to be qualitatively captured in the minimal model accurately. In this model the solvent condition is portrayed through attractive LJ interactions between identical bead types and WCA repulsive interactions between different bead types. Aggregation of the POSS cubes is observed in the AA and CG simulations that each account for POSS interactions with hexane solvent molecules. Clustering of the POSS cubes with tethers is not observed in these simulations. Both of these behaviors were also observed in previous simulations employing the minimal model.

VI. CONCLUSIONS

A CG model and force field were developed for capturing self-assembly of nonyl-tethered POSS molecules dissolved in hexane solvent. The model consists of a rigid cube with eight bead interaction sites on the vertices and a flexible tether consisting of bonded beads. A one-to-one mapping of groups of atoms in the atomistic molecule onto CG beads in the mesoscale model was established. The CG force field consists of solvent-mediated effective potentials that were derived using a structural-based coarse-graining numerical iteration scheme. It was demonstrated that these potentials could reproduce closely target intermolecular RDFs between interaction sites on bare silsesquioxane cages, target intramolecular bond length probability distributions, and target intramolecular bond angle probability distributions that were computed from the underlying explicit atom simulations. Furthermore, the potentials reproduced fairly well target RDFs for cube centers of mass, which are not interaction sites specified in the model.

The CG force field was evaluated in several ways. First, we found that different initial guesses for the effective potentials could reproduce a similar effective pair potential for bare POSS cages via Eq. (1) at low temperature but not at higher temperature. The differences in the forms of the resulting potentials in the latter case demonstrate the nonuniqueness inherent in the approach undertaken here. The accuracy of the potentials was evaluated by computing merit functions that measure the level of agreement between the target atomistic and CG probability distribution functions. Second, a different iteration algorithm was proposed that could reproduce the same overall effective potential for bare cube corner bead sites using the same WCA initial guess for the potential. This algorithm required a smaller number of iteration steps to achieve convergence compared to Eq. (1), and thus appears to be more efficient in deriving CG potentials for the system studied here.

The self-assembly behavior in both the atomistic and CG simulated systems was compared to validate the accuracy of the CG model. In both simulations small aggregates containing a few POSS molecules were observed. Several types of local packing of the silsesquioxane cages and various tether conformations were observed in the atomistic simulations, and these features were captured in the CG systems. Both the atomistic and CG simulations of monotethered POSS molecules resulted in self-assembly of the molecules into small aggregates with the cages comprising the aggregate cores and the tethers pointing outward. The CG model afforded a savings in computing time of roughly two orders of magnitude. It can be utilized next in simulations to predict the types of self-assembled structures or phases arising in systems containing thousands of nonyl-tethered POSS molecules dissolved in hexane. These structures can then be compared with the corresponding experimental studies.

Comparisons between the CG monotethered POSS model developed here and the corresponding minimal model developed earlier[2] indicate that the intermolecular interactions between cages are long ranged. Short-ranged interactions favor face-face intermolecular packing of cubes and longer-ranged interactions are necessary to reproduce other types of intermolecular orientations. The effective bond bending potentials in the CG model are important for capturing both the orientation of the tether relative to the attached silsesquioxane cage and the bond angles along the tether. These features are not captured faithfully by the fully flexible tether in the minimal model. The neutral solvent condition is qualitatively captured appropriately in the minimal model, as aggregation of POSS cubes, and not of POSS cubes with tethers, is observed in the AA, CG, and minimal model simulations. The CG model developed here incorporates the appropriate intermolecular packings between POSS cages and intramolecular orientation of the tether relative to the cube. These elements, which are absent in the minimal model, should be important for accurately predicting the types of self-assembled mesoscopic structures that arise in solution, and for yielding accurate simulation results that can be compared directly with experiments to better understand the self-assembly processes involved.

ACKNOWLEDGMENTS

The authors thank the National Science Foundation for financial support under Grant No. DMR-0103399. One of the authors (E.R.C.) is grateful to S. K. Kandasamy and Z. L. Zhang for fruitful discussions. The University of Michigan Center for Advanced Computing is acknowledged for computer cluster support. The all-atom molecular dynamics simulations were primarily performed on the VAMPIRE cluster at Vanderbilt University.

APPENDIX: RELATIONSHIP BETWEEN RADIAL DISTRIBUTION FUNCTIONS AND POTENTIALS OF MEAN FORCE FOR MOLECULAR SOLUTES

One commonly sees in the published literature references to potentials of mean force obtained from site-site radial distribution functions. This is particularly so in methods used to refine neutron scattering results[13] and to develop coarse-grained interaction potentials from detailed atomistic simulations.[13,15,16] However, many of the statements concerning potentials of mean force are incorrect, as we demonstrate below.

For spherically symmetric pure fluids, the potential of mean force (PMF) $W(r)$ between two of the molecules in the fluid is well defined and given by (for example, see Gray and Gubbins[31])

$$W(r) = -k_B T \ln[g(r)], \tag{A1}$$

where k_B is Boltzmann's constant, T is the absolute temperature, and $g(r)$ is the radial distribution function between the molecules in the fluid. The PMF is state dependent (a function of both density ρ and temperature T). Strictly speaking, since $W(r)$ is a function of density and temperature, it more appropriately should be written as $W(\rho,T;r)$. It satisfies the exact limiting relation[31]

$$\lim_{\rho\to 0} W(\rho,T;r) \to u(r), \tag{A2}$$

where $u(r)$ is the (spherically symmetric) pair potential between the molecules in the fluid. Additionally, $W(\rho,T;r)$ has the significance of being an effective potential, in the sense that a simulation of two molecules performed with $W(r)$ as the potential will yield the same structure as the many-body simulation at density ρ and temperature T. This is because for a two-body system the pair distribution function interacting with potential $W(\rho,T;r)$ will be given by $e^{-W(\rho,T;r)/k_B T}$, which by definition is the structure of the many-body fluid at density ρ and temperature T.

For molecular fluids, there is a similar set of results relating the PMF $W(12)$ to the pair distribution function $g(12)$. In these functions, numbers 1 and 2 stand, respectively, for the multidimensional vectors $\mathbf{r}_1\mathbf{\Omega}_1$ and $\mathbf{r}_2\mathbf{\Omega}_2$, where $\mathbf{r}_i$ is the position of the center of mass of molecule i and $\mathbf{\Omega}_i$ is the vector specifying the orientation of molecule i in space. (For example, for linear molecules, $\mathbf{\Omega}_i$ is a vector of two elements, and for nonlinear molecules contains the three Euler angles.) In this case, $W(12)$ is well defined and given by

$$W(12) = -k_B T \ln[g(12)]. \tag{A3}$$

As in the case of simple fluids, the PMF is state dependent (a function of both density ρ and temperature T) and satisfies the exact limiting relation

$$W(12) \to u(12) \quad \text{as } \rho \to 0, \tag{A4}$$

where $u(12)$ is the anisotropic pair potential between the molecules in the fluid. As before, strictly speaking, $W(12)$ is a function of density and temperature, so it more appropriately should be written as $W(\rho,T;12)$. Again, as in the case of monatomic fluids, $W(\rho,T;12)$ has the significance of being an effective potential, in the sense that a simulation of two molecules performed with $W(\rho,T;12)$ as the potential will yield the same structure as the many-body simulation at density ρ and temperature T, since for a two-body system the pair distribution function interacting with anisotropic potential $W(\rho,T;12)$ will be given by $e^{-W(\rho,T;12)/k_B T}$, which by definition is the structure of the many-body fluid at density ρ and temperature T.

Now let us consider site-site potentials and site-site pair distribution functions. For two identical m-site molecules that interact via site-site potentials, the total potential $u(12)$ is given by[31]

$$u(12) = \sum_{\alpha=1}^{m} \sum_{\beta=1}^{m} u_{\alpha\beta}(r_{12}^{\alpha\beta}), \tag{A5}$$

where $u_{\alpha\beta}(r)$ is the site-site potential between site α in one molecule and site β in another molecule, and $r_{12}^{\alpha\beta} = |\mathbf{r}_{12}^{\alpha\beta}| = |\mathbf{r}_1^{\alpha} - \mathbf{r}_2^{\beta}|$ where $\mathbf{r}_i^{\gamma}$ is the position of site γ in molecule i; hence, $r_{12}^{\alpha\beta}$ is the distance between site α in molecule 1 and site β in molecule 2. Although the site-site potentials are spherically symmetric, the total potential $u(12)$ is evidently not spherically symmetric since changing relative orientations of the two molecules changes many or all of the $r_{12}^{\alpha\beta}$. Note that the decomposition of $u(12)$ into site-site interaction potentials is an approximation, and is not exact from the point of view of quantum mechanics.

The site-site pair (or radial) distribution functions, $g_{\alpha\beta}(r)$, are exact quantities defined to be somewhat analogous to the radial distribution function of simple fluids, but with different properties. Specifically, $g_{\alpha\beta}(r)$ is defined to be proportional to the probability of finding site α in molecule 1 and site β in molecule 2 separated by distance r independent of the orientation of the two molecules. Thus, it has a precise definition in terms of the molecular distribution function $g(12)$ as follows:[31]

$$g_{\alpha\beta}(r) = g_{\alpha\beta}(|\mathbf{r}' - \mathbf{r}''|) = \frac{1}{\Omega^2} \int\int \delta(\mathbf{r}_1^{\alpha} - \mathbf{r}')\,\delta(\mathbf{r}_2^{\beta} - \mathbf{r}'')\, g(12)\, d\mathbf{\Omega}_1 d\mathbf{\Omega}_2, \tag{A6}$$

where $\Omega = \int d\mathbf{\Omega}_i = 4\pi$ for a linear molecule and $8\pi^2$ for a nonlinear molecule. A molecule does not need to interact via a potential made up of site-site interactions in order to obtain $g_{\alpha\beta}(r)$—the site-site radial distribution functions can be obtained independent of the assumption of site-site interactions. This is clear because $g_{\alpha\beta}(r)$ can be obtained experimentally from neutron scattering, where the interactions between the molecules are, in general, unknown exactly, and are not decomposable into site-site interactions. Thus, the site-site radial distribution function (SSRDF) is a much more fundamental quantity than a site-site potential (SSP).

Let us now consider the quantity frequently cited in publications as the site-site analog of the PMF,[13,15,16] viz.,

$$W_{\alpha\beta}(r) = -k_B T \ln[g_{\alpha\beta}(r)]. \tag{A7}$$

We shall demonstrate that this quantity has no asymptotic properties similar to those of the spherically symmetric PMF $W(\rho,T;r)$ for simple fluids or the molecular PMF

$W(\rho,T;12)$ for molecular fluids. Consider Eq. (A7) in the limit as $\rho\to 0$. Then

$$\begin{aligned} W_{\alpha\beta}(r) &\to -k_BT\ln[\lim_{\rho\to 0} g_{\alpha\beta}(r)] \\ &= -k_BT\ln\left[\lim_{\rho\to 0}\frac{1}{\Omega^2}\int\int \delta(\mathbf{r}_1^{\alpha}-\mathbf{r}') \right. \\ &\quad \left. \times\delta(\mathbf{r}_2^{\beta}-\mathbf{r}'')g(12)d\mathbf{\Omega}_1 d\mathbf{\Omega}_2\right] \\ &= -k_BT\ln\left[\lim_{\rho\to 0}\frac{1}{\Omega^2}\int\int \delta(\mathbf{r}_1^{\alpha}-\mathbf{r}') \right. \\ &\quad \left. \times\delta(\mathbf{r}_2^{\beta}-\mathbf{r}'')e^{-u(12)/k_BT}d\mathbf{\Omega}_1 d\mathbf{\Omega}_2\right]. \end{aligned} \tag{A8}$$

Clearly,

$$\lim_{\rho\to 0} W_{\alpha\beta}(r) \neq u_{\alpha\beta}(r). \tag{A9}$$

Hence, in determining the effective site-site potential either for neutron scattering analysis or for coarse graining (as is the case in the present paper), there is no fundamentally sound reason for choosing as the initial guess

$$u^0_{\alpha\beta}(r) = -k_BT\ln[g_{\alpha\beta}(r)]. \tag{A10}$$

Clearly, there is also no theoretical justification for the iterative scheme based on the assumption that $W_{\alpha\beta}(r)$ has any PMF-like relationship to $g_{\alpha\beta}(r)$. Thus, the iteration scheme

$$u^{(n+1)}_{\alpha\beta}(r) = u^{(n)}_{\alpha\beta}(r) - k_BT\ln\left[\frac{g^{(n)}_{\alpha\beta}(r)}{g^{\text{target}}_{\alpha\beta}(r)}\right] \tag{A11}$$

also has no fundamental basis in theory. However, it does have the correct stationary point [i.e., $u^{(n)}_{\alpha\beta}(r)$ must converge when $g^{(n)}_{\alpha\beta}(r)=g^{\text{target}}_{\alpha\beta}(r)$, where $g^{\text{target}}_{\alpha\beta}(r)$ could either be experimental data (as in the neutron scattering application) or the site-site radial distribution function from a more atomistically detailed simulation (as in the coarse-graining application)]. However, as we have demonstrated in this paper, other iteration schemes that begin from a simple initial guess [different from Eq. (A10)] and use a different iteration scheme from Eq. (A11) (but having the same convergence point) will yield the same effective site-site interaction.

Another quantity frequently spoken of in PMF terms is related to the center-to-center (CC) distribution function given by

$$\begin{aligned} g_{\text{CC}}(r) &= g_{\text{CC}}(|\mathbf{r}'-\mathbf{r}''|) \\ &= \frac{1}{\Omega^2}\int\int \delta(\mathbf{r}_1^{C}-\mathbf{r}') \\ &\quad \times\delta(\mathbf{r}_2^{C}-\mathbf{r}'')g(12)d\mathbf{\Omega}_1 d\mathbf{\Omega}_2 \end{aligned} \tag{A12}$$

and the related PMF-like quantity

$$W_{\text{CC}}(\rho,T;r) = -k_BT\ln[g_{\text{CC}}(r)]. \tag{A13}$$

Note that in Eq. (A12) the CC distribution function is obtained by fixing the relative positions of the centers of the two molecules, then integrating over all orientations of the two molecules. In this case, $W_{\text{CC}}(\rho,T;r)$ has the significance that a simulation of two *spherically symmetric* molecules performed with interaction potential $W_{\text{CC}}(\rho,T;r)$ would have the same center-to-center structure as the many-body molecular fluids. It has the limiting property that

$$\begin{aligned} \lim_{\rho\to 0} W_{\text{CC}}(\rho,T;r) &= -k_BT\ln[\lim_{\rho\to 0} g_{\text{CC}}(r)] \\ &= -k_BT\ln\left[\frac{1}{\Omega^2}\int\int \delta(\mathbf{r}_1^{C}-\mathbf{r}') \right. \\ &\quad \left. \times\delta(\mathbf{r}_2^{C}-\mathbf{r}'')\lim_{\rho\to 0} g(12)d\mathbf{\Omega}_1 d\mathbf{\Omega}_2\right]_1 \\ &= -k_BT\ln\left[\frac{1}{\Omega^2}\int\int \delta(\mathbf{r}_1^{C}-\mathbf{r}') \right. \\ &\quad \left. \times\delta(\mathbf{r}_2^{C}-\mathbf{r}'')e^{-u(12)/k_BT}d\mathbf{\Omega}_1 d\mathbf{\Omega}_2\right]. \end{aligned} \tag{A14}$$

Thus, if we define the spherically averaged potential between two molecules as

$$\begin{aligned} u_{\text{CC}}(\rho,T;r) &= \frac{1}{\Omega^2}\int\int \delta(\mathbf{r}_1^{C}-\mathbf{r}') \\ &\quad \times\delta(\mathbf{r}_2^{C}-\mathbf{r}'')u(12)d\mathbf{\Omega}_1 d\mathbf{\Omega}_2, \end{aligned} \tag{A15}$$

then clearly $\lim_{\rho\to 0} W_{\text{CC}}(\rho,T;r)\neq u_{\text{CC}}(r)$; rather, in the limit of low density $W_{\text{CC}}(\rho,T;r)$ becomes what is known as the reference-averaged Mayer f-function (RAM) potential between the centers of the molecule used in the RAM perturbation theory of molecular fluids.[32]

[1] R. Knischka, F. Dietsche, R. Hanselmann, H. Frey, R. Mulhaupt, and P. J. Lutz, Langmuir **15**, 4752 (1999); B. S. Kim and P. T. Mather, Macromolecules **35**, 8378 (2002); K. M. Kim, D. K. Keum, and Y. Chujo, *ibid.* **36**, 867 (2003); G. Cardoen and E. B. Coughlin, *ibid.* **37**, 5123 (2004).

[2] E. R. Chan, X. Zhang, C.-Y. Lee, M. Neurock, and S. C. Glotzer, Macromolecules **38**, 6168 (2005).

[3] X. Zhang, E. R. Chan, and S. C. Glotzer, J. Chem. Phys. **123**, 184718 (2005); E. R. Chan, L. C. Ho, and S. C. Glotzer, *ibid.* **125**, 064905 (2006).

[4] T. C. Ionescu, F. Qi, C. McCabe, A. Striolo, J. Kieffer, and P. T. Cummings, J. Phys. Chem. B **110**, 2502 (2006).

[5] A. Striolo, C. McCabe, and P. T. Cummings, Macromolecules **38**, 8950 (2005).

[6] A. Striolo, C. McCabe, and P. T. Cummings, J. Phys. Chem. B **109**, 14300 (2005).

[7] A. Striolo, C. McCabe, E. R. Chan, S. C. Glotzer, and P. T. Cummings, J. Phys. Chem. B (in press).

[8] J. Baschnagel, K. Binder, P. Doruker *et al.*, Adv. Polym. Sci. **152**, 41 (2000); K. Kremer and F. Muller-Plathe, MRS Bull. **26**, 205 (2001); Mol. Simul. **28**, 729 (2002); F. Muller-Plathe, Soft Mater. **1**, 1 (2003); S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein, J. Phys.: Condens. Matter **16**, R481 (2004); G. Lu and E. Kaxiras, in *Handbook of Theoretical and Computational Nanotechnology* edited by M. Rieth and W. Schommers (American Scientific Publishers, Stevenson Ranch, CA, 2005), Vol. 10, p. 1.

[9] S. C. Glotzer and W. Paul, Annu. Rev. Mater. Res. **32**, 401 (2002).

[10] F. Muller-Plathe, ChemPhysChem **3**, 754 (2002).

[11] M. J. Stevens, J. Chem. Phys. **121**, 11942 (2004).

[12] R. Faller, H. Schmitz, O. Biermann, and F. Muller-Plathe, J. Comput. Chem. **20**, 1009 (1999); H. Meyer, O. Biermann, R. Faller, D. Reith, and F. Muller-Plathe, J. Chem. Phys. **113**, 6264 (2000); D. Reith, H. Meyer, and F. Muller-Plathe, Macromolecules **34**, 2335 (2001); D. Reith, H. Meyer, and F. Muller-Plathe, Comput. Phys. Commun. **148**, 299 (2002).

[13] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995); A. K. Soper, Chem. Phys. **202**, 295 (1996).
[14] W. Tschop, K. Kremer, J. Batoulis, T. Burger, and O. Hahn, Acta Polym. **49**, 61 (1998); J. Eilhard, A. Zirkel, W. Tschop, O. Hahn, K. Kremer, O. Scharpf, D. Richter, and U. Buchenau, J. Chem. Phys. **110**, 1819 (1999); A. P. Lyubartsev, M. Karttunen, I. Vattulainen, and A. Laaksonen, Soft Mater. **1**, 121 (2003); A. P. Lyubartsev, Eur. Biophys. J. Biophys. Lett. **35**, 53 (2005).
[15] D. Reith, M. Putz, and F. Muller-Plathe, J. Comput. Chem. **24**, 1624 (2003).
[16] H. S. Ashbaugh, H. A. Patel, S. K. Kumar, and S. Garde, J. Chem. Phys. **122**, 104908 (2005).
[17] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, J. Chem. Phys. **120**, 10896 (2004); S. Izvekov and G. A. Voth, J. Phys. Chem. B **109**, 2469 (2005); S. Izvekov, A. Violi, and G. A. Voth, *ibid.* **109**, 17019 (2005).
[18] A. Violi and G. A. Voth, in *High Performance Computing and Communications, Proceedings Lecture Notes in Computer Science*, edited by L. T. Yang, O. F. Rana, B. Di Martino, and J. Dongarra (Springer-Verlag, Berlin, 2005), Vol. 3726, p. 938.
[19] R. D. Groot and P. B. Warren, J. Chem. Phys. **107**, 4423 (1997); M. Murat and K. Kremer, *ibid.* **108**, 4340 (1998); S. D. Chao, J. D. Kress, and A. Redondo, *ibid.* **122**, 234912 (2005); R. L. C. Akkermans and W. J. Briels, *ibid.* **114**, 1020 (2001); W. J. Briels and R. L. C. Akkermans, Mol. Simul. **28**, 145 (2002); P. G. Bolhuis, A. A. Louis, J. P. Hansen, and E. J. Meijer, J. Chem. Phys. **114**, 4296 (2001); P. G. Bolhuis and A. A. Louis, Macromolecules **35**, 1860 (2002); C. N. Likos, N. Hoffmann, H. Lowen, and A. A. Louis, J. Phys.: Condens. Matter **14**, 7681 (2002).
[20] J. Baschnagel, K. Binder, W. Paul, M. Laso, U. W. Suter, I. Batoulis, W. Jilge, and T. Burger, J. Chem. Phys. **95**, 6014 (1991); S. J. Marrink, A. H. de Vries, and A. E. Mark, J. Phys. Chem. B **108**, 750 (2004); P. K. Depa and J. K. Maranas, J. Chem. Phys. **123**, 094901 (2005).
[21] R. L. Henderson, Phys. Lett. **49A**, 197 (1974).
[22] D. A. McQuarrie, *Statistical Mechanics* (University Science Books, Sausalito, CA, 2000).
[23] A. R. Leach, *Molecular Modeling: Principles and Applications*, 2nd ed. (Prentice-Hall, Harlow, 2001).
[24] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
[25] W. F. van Gunsteren, H. J. C. Berendsen, and J. A. C. Rullmann, Mol. Phys. **44**, 69 (1981); G. S. Grest, M. D. Lacasse, K. Kremer, and A. M. Gupta, J. Chem. Phys. **105**, 10583 (1996).
[26] W. Smith and T. R. Forester, J. Mol. Graphics **14**, 136 (1996).
[27] A. L. Frischknecht and J. G. Curro, Macromolecules **36**, 2122 (2003).
[28] M. G. Martin and J. I. Siepmann, J. Phys. Chem. B **102**, 2569 (1998).
[29] H.-C. Li, C.-Y. Lee, C. McCabe, A. Striolo, and M. Neurock, J. Phys. Chem. A **111**, 3577 (2007).
[30] J. A. Barker and D. Henderson, Mol. Phys. **21**, 187 (1971).
[31] C. G. Gray and K. E. Gubbins, *Theory of Molecular Fluids* (Oxford University Press, Oxford, 1985).
[32] I. Nezbeda and W. R. Smith, Chem. Phys. Lett. **81**, 79 (1981); W. R. Smith, I. Nezbeda, T. W. Melnyk, and D. D. Fitts, Faraday Discuss. Chem. Soc. **66**, 130 (1978); I. Nezbeda and W. R. Smith, Chem. Phys. Lett. **64**, 146 (1979); J. W. Perram and L. R. White, Mol. Phys. **28**, 527 (1974).