8-11-2011

Reintroducing explicit solvent to a solvent-free coarse-grained model

Seung Ha Kim
Iowa State University

Monica H. Lamm
Iowa State University, mhlamm@iastate.edu

Reintroducing explicit solvent to a solvent-free coarse-grained model

Abstract

A unique coarse-grained modeling scheme that combines a systematic, solvent-free multiscale coarse-graining algorithm for a complex macromolecule with an existing coarse-grained solvent model is proposed. We show that this procedure efficiently and reliably describes the interactions for complex macromolecules, using the specific example of dendrimers binding phenanthrenes in water. The experimentally measured binding capacity is predicted by the unique coarse-grained modeling approach; the conditions for this simulation are beyond what could be reasonably simulated with an all-atom molecular dynamics simulation.

Disciplines

Biochemical and Biomolecular Engineering | Organic Chemistry | Polymer Chemistry

Comments

This article is published as Kim, Seung Ha, and Monica H. Lamm. "Reintroducing explicit solvent to a solvent-free coarse-grained model." Physical Review E 84, no. 2 (2011): 025701. doi:10.1103/PhysRevE.84.025701. Posted with permission.

PHYSICAL REVIEW E **84**, 025701(R) (2011)

Reintroducing explicit solvent to a solvent-free coarse-grained model

Seung Ha Kim and Monica H. Lamm*
Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa 50011, USA
(Received 31 January 2011; revised manuscript received 30 June 2011; published 11 August 2011)

A unique coarse-grained modeling scheme that combines a systematic, solvent-free multiscale coarse-graining algorithm for a complex macromolecule with an existing coarse-grained solvent model is proposed. We show that this procedure efficiently and reliably describes the interactions for complex macromolecules, using the specific example of dendrimers binding phenanthrenes in water. The experimentally measured binding capacity is predicted by the unique coarse-grained modeling approach; the conditions for this simulation are beyond what could be reasonably simulated with an all-atom molecular dynamics simulation.

DOI: 10.1103/PhysRevE.84.025701 PACS number(s): 02.70.−c, 82.35.−x, 87.15.A−, 87.85.Qr

Coarse-grained (CG) molecular models have broad appeal for simulation-based investigations of complex macromolecules, such as polymers, surfactants, proteins, and lipid membranes, because they permit interrogation of these systems at length and time scales much larger than is possible with all-atom molecular models. Most CG approaches can be classified into one of two categories. In the first category, termed indirect parametrization, are methods in which the potential parameters of a preselected analytical form are optimized by calibration against thermodynamic or structural properties. An example is the MARTINI force field for biological molecules [1], whose parameters are based on oil-water partitioning coefficients. In the second category, called direct parametrization, the CG potentials are determined from an explicit atom molecular dynamics (MD) simulation. One example is the multiscale coarse-graining (MS-CG) method, which derives CG parameters from force matching [2]. In the former category, a force field such as MARTINI can be easily applied whenever the target system is changed with little to no reparametrization required. Unfortunately, because the CG parameters are not directly based on the underlying atomistic forces, it is difficult to reproduce accurate local ($<$10 Å) structural details for a specific system. For example, the MARTINI force field is incapable of predicting the preferred positions for noncovalent binding between a flexible macromolecule and a small organic guest molecule [3]. In the latter category, the MS-CG method has the advantage of being systematic; the CG force field is evaluated from data collected along the trajectories of reference atomistic MD simulations.

The MS-CG method has been implemented for a solvated lipid bilayer with up to 12 CG sites [4]. However, storage and memory requirements make it computationally difficult to derive CG potentials for complex biological systems with greater than 15 defined CG site interaction types [2]. An alternative approach, called the solvent-free MS-CG model, derives effective CG potentials between sites on the solute molecules while integrating out the explicit representation of the solvent molecules [5]. This approach was recently implemented for a lipid bilayer [5] and for polyglutamine peptides [6]. For relatively rigid molecules, such as lipids, the local and long-range structure calculated from the solvent-free CG MD and the reference atomistic MD simulations are nearly identical [5]. For flexible molecules, the absence of an explicit solvent can cause the solvent-free MS-CG MD simulations to produce an increased tendency toward intra- and intermolecular aggregation. This has been observed in the solvent-free CG model for polyglutamine; further evidence for this solvent-free CG effect is provided in the Supplemental Material [7]. Specifically, the radial distribution functions obtained for the flexible molecules in the solvent-free CG model indicate structures that are highly ordered, compared to the explicit solvent CG model. The drastic reduction in degrees of freedom for the solvent-free CG model leads to configurational entropy loss. The loss of configurational entropy upon coarse graining has been quantified for hydrocarbon chains and is shown to increase as the flexibility of the chain increases [8]. Thus, while solvent-free CG models may work well for lipids, an explicit solvent CG model is required for problems concerning flexible macromolecules in solution.

In this Rapid Communication, we present a method that combines the computational efficiency of deriving solvent-free MS-CG potentials from force matching with the improved reliability of retaining the solvent degrees of freedom (i.e., reducing configurational entropy loss) in the CG MD simulation by using independently derived CG solvent potentials. We illustrate the accuracy and convenience of this approach by modeling a mixture of generation 5 poly(amidoamine) (G5-PAMAM) dendrimers and phenanthrene (Phe) in water to calculate the binding properties of PAMAM. Solvent-free CG potentials for PAMAM and Phe in water at neutral *p*H were obtained by the MS-CG approach described elsewhere [5,9,10]. A brief summary follows. The first step of the MS-CG procedure is to obtain the atomic positions and forces sampled from an equilibrated, atomistic MD simulation. The degrees of freedom in the atomic configurations are reduced by mapping groups of atoms to defined CG sites and computing the net forces acting on the coarse-grained sites. If $\mathbf{f}_{ij}(\mathbf{r}_i,\mathbf{r}_j)$ is the nonbonded CG force acting on the ith CG site due to the jth CG site, and it is assumed to depend linearly on m unknown parameters $p_1, p_2, \ldots, p_m$, the CG pair force is expressed by $\mathbf{f}_{ij}(\mathbf{r}_i,\mathbf{r}_j,p_1,p_2,\ldots,p_m)$. To obtain the CG potential in a systematic way, cubic spline or B-spline functions are fitted to the CG forces to enable a smooth curvature across mesh points. Specifically, B-spline functions improve the force-matching performance because they can reduce the memory requirement and increase the accuracy [2]. Then, the m

*mhlamm@iastate.edu

unknown parameters are optimized based on a least-squares method, which minimizes the difference between the net forces and the fitted pair forces and is expressed as

$$\chi^2 = \sum_{i=1}^{N_i} \left| \mathbf{F}_i^{\text{atomistic}} - \mathbf{F}_i^{\text{predicted}}(\mathbf{r}_i, p_1, p_2, \ldots, p_m) \right|^2, \quad (1)$$

where $\mathbf{F}_i^{\text{atomistic}}$ and $\mathbf{F}_i^{\text{predicted}}$ are the reference atomistic force field and the calculated force field, respectively. The parameters obtained in this way from each configuration are averaged over the total number of atomic configurations sampled [2]. In this work, the solvent-free potentials were calculated under the solvent-free condition of the force-matching method, in which the solute molecules only are considered during the reference atomistic MD simulation to get the CG pair forces, and, thus, the solvent effect is implicitly included in the CG potentials of the solute molecules [5]. The reference atomistic simulation contained 27 G5-PAMAM dendrimers and 216 Phe molecules with 1 389 280 explicit water molecules and 3456 Cl^- counterions at neutral *p*H. A detailed description of the all-atom simulation methodology is provided in the Supplemental Material [7].

For the CG model, the atoms in the dendrimers and the Phe molecules have been grouped into four CG sites as shown in Fig. 1. The CG mapping scheme used by Ref. [11] was applied; this scheme is based on the MARTINI CG force field [1]. To determine the bonded interactions in the CG system, a harmonic potential was applied: $V(r) = \frac{1}{2}k_r(r - r_0)^2$ and $V(\theta) = \frac{1}{2}k_\theta(\theta - \theta_0)^2$, where $V(r)$ and $V(\theta)$ are bond and angle potentials, and k_r and k_θ are the bond force constant and the angle force constant, respectively. The parameters for the bonded potentials were determined by inverse Boltzmann fitting of the bond distributions [12]. The bonded interactions

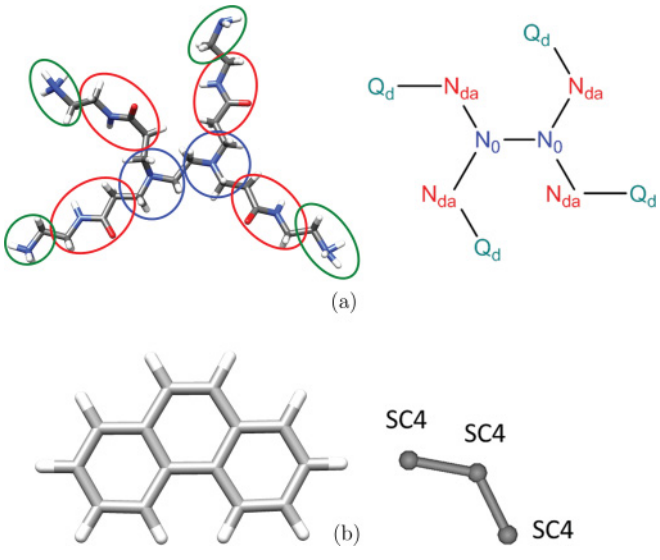

FIG. 1. (Color online) Atomistic (left-hand side) and coarse-grained (right-hand side) representation. In the atomistic structure, gray spheres represent carbon; blue, nitrogen; red, oxygen; and white, hydrogen. (a) PAMAM (for clarity, only the core is shown). For coarse graining, blue circles of the atoms map into N_0; red, N_{da}; and green, Q_d. (b) Phenanthrene. Each aromatic ring is represented by a single CG site, SC_4.

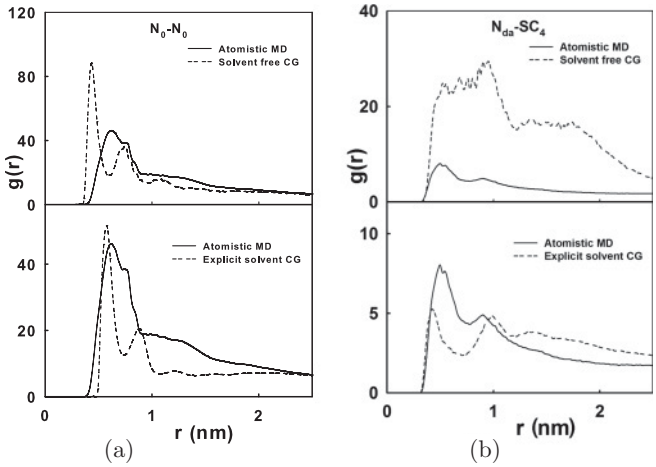

FIG. 2. Radial distribution functions (RDFs) between CG types of PAMAM dendrimers and Phe from atomistic (solid line) and coarse-grained (dashed line) molecular dynamics simulations. The dashed lines in the upper figures represent the solvent-free CG model, and in the bottom figures the explicit solvent CG model. (a) The site-site RDF for N_0-N_0, and (b) the site-site RDF for N_{da}-SC_4. For clarity in comparison, the distributions are only shown up to 2.5 nm. All distributions shown approach unity at large r (not shown).

are calculated by the simple Boltzmann probability $V(r) = -k_B T \ln[P(r)]$, where $P(r)$ is calculated by the bonded interaction distribution of the all-atom simulations. For the bond angle potentials, the initial guess of the angle force constants k_θ are selected by the inverse Boltzmann fitting method and determined in an iterative way because of the flexibility of the dendrimer; the equilibrium angles are obtained from the angle distributions. The nonbonded and bonded force parameters for the G5-PAMAM dendrimers and Phe molecules are given in the Supplemental Material [7].

In order to validate the solvent-free CG potentials, the site-site radial distribution functions (RDF) between solvent-free CG sites from G5-PAMAM and Phe are compared with those from the atomistic MD simulations (the top row of Fig. 2). N_0 and N_{da} are the CG atom types of G5-PAMAM dendrimers, and SC_4 is the CG atom type for Phe. The RDF of N_0-N_0 shows the structures of the dendrimers, and the RDF of N_{da}-SC_4 indicates the strength of the interactions between PAMAM dendrimers and phenanthrene molecules. The intensities of the RDFs from the solvent-free simulation are higher than in the atomistic simulation, indicating that the dendrimers have aggregated and that the interaction of the dendrimers and Phe molecules in the CG simulations are much stronger than in the all-atom simulations. This tendency of the CG model to overpredict aggregation occurs because of the absence of explicit water molecules and because the dendrimer is a flexible macromolecule. Data provided in the Supplemental Material shows that this enhanced aggregation occurs because of the absence of explicit CG water molecules in the system, and is not due to the coarse-graining scheme itself [7].

To overcome this limitation, we propose a methodology (an explicit solvent CG model), which introduces the CG water molecules to the formerly solvent-free system of G5-PAMAM and Phe. The MARTINI water model was selected because it is compatible with the CG mapping scheme applied to the

solutes. Even though the degrees of freedom of the system increase after introducing the CG solvent molecules, the explicit solvent CG model (109 944 atoms) is still computationally efficient relative to the atomistic simulation (4 292 364).

The CG pair potentials for interactions involving water have the Lennard-Jones form $U_{\mathrm{LJ}}(r) = \varepsilon_{ij}[(\frac{\sigma_{ij}}{r})^{12} - 100(\frac{\sigma_{ij}}{r})^{6}]$, where ε_{ij} is the strength of the interaction, and σ_{ij} represents the closest distance between ith and jth CG sites. Because the solvent-free CG potentials already include implicit water effects, the strength of the solute-water interaction parameters had to be rationalized based on the physical properties of CG types, such as polarity. This was accomplished as follows. The CG solute sites were divided into two types: hydrophilic (N_0, N_{da}, Q_d) and hydrophobic (SC_4). The strengths of the interaction parameters were modified such that $\varepsilon_{\mathrm{hydrophobic}}(8\ \mathrm{kJ\,mol^{-1}}) < \varepsilon_{\mathrm{hydrophilic}}(15\ \mathrm{kJ\,mol^{-1}}) < \varepsilon_{\mathrm{water}}(20\ \mathrm{kJ\,mol^{-1}})$, where $\varepsilon_{\mathrm{water}}$ is the same value used in the MARTINI CG water model. CG MD simulations in the NVT ensemble were run after introducing explicit CG water modeled with the modified MARTINI potential. The RDFs between dendrimer and Phe sites are shown in the bottom row of Fig. 2. The CG system with explicit water more closely matches the RDFs observed from the reference atomistic MD simulations.

To test the concentration dependence of the CG potentials derived with the explicit solvent CG approach, the number of PAMAM dendrimers was varied from four to 27 while keeping a fixed number (216) of Phe molecules. Figure 3 shows the fraction of Phe bound to PAMAM in 216 Phe molecules, computed as the accumulated probability density function of Phe from the core to the surface of the dendrimer. As a consequence, Fig. 3 indicates the capacity of the dendrimer to encapsulate Phe molecules. As mentioned above, the CG potentials for all of the CG simulations in Fig. 3 were derived from the atomistic simulation with 27 PAMAM dendrimers and 216 Phe molecules in explicit water. Since the intensity of fraction of bound Phe molecules in the solvent-free CG model is higher compared to the atomistic and the explicit solvent CG

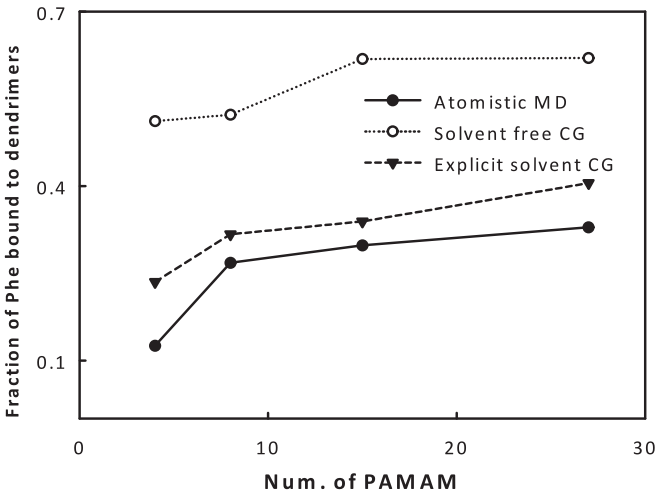

FIG. 3. Fraction of Phe bound to the dendrimer. The number of Phe molecules is fixed at 216. Atomistic MD is shown by closed circles, the solvent-free CG model by open circles, and the explicit solvent CG model by closed triangles.

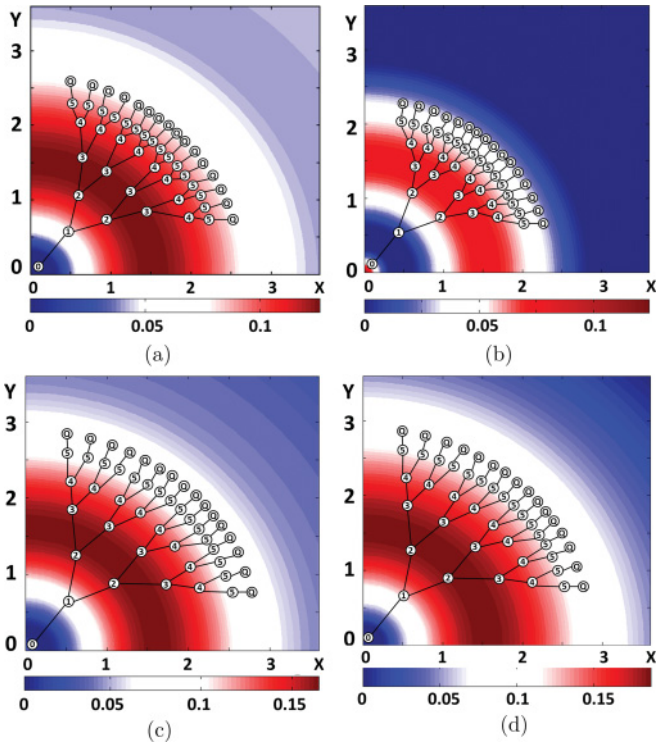

FIG. 4. (Color online) Two-dimensional probability distribution function $P(r)$ of Phe molecules from the core of each dendrimer. The white circles indicate the branch points of the dendrimers. The numbers in the circles represent the generation of the dendrimer, and Q is the abbreviation for Q_d. The color bars express the intensity of $P(r)$. The unit of the X and Y axes is nm. (a) Atomistic simulation, (b) solvent-free CG, (c) explicit solvent CG (27 dendrimers), and (d) explicit solvent CG (216 dendrimers).

model, the strength of the binding interaction of Phe molecules is overestimated. Therefore the solvent-free CG model does not reproduce the interactions of PAMAM dendrimers with Phe, and the explicit solvent CG model transfers well to different relative concentrations.

Earlier experimental measurements on the fluorescence resonance energy transfer (FRET) between G5-PAMAM and Phe showed that FRET efficiency increased as the molar ratio of PAMAM:Phe increased from 1:10 to 1:1 [13]. To investigate the effect of molar ratio on binding capacity in the model system, we increased the number of G5-PAMAM dendrimers from 27 to 216, while keeping the number of Phe fixed at 216. Figure 4 shows the two-dimensional probability distributions function, $P(r)$, of Phe molecules from the core of each dendrimer, overlayed with the average location of each dendrimer branch point. Comparing the solvent-free case to the reference atomistic simulation shows that the peak intensity of the solvent-free model is nearly four times higher. The peak intensity of the CG model with explicit water is much closer to the reference atomistic MD system. In the simulation with 216 dendrimers, peak intensity has increased, consistent with the molar ratio dependency effect on binding capacity observed by the experiment performed in Ref. [13]. We note that a simulation of this scale is not feasible with atomistic MD.

In this Rapid Communication we have introduced an approach for restoring the configurational entropy lost when

the solvent degrees of freedom are removed from a system that contains flexible macromolecules in solution. In this procedure, the effective solvent-free solute-solute potentials are derived by MS-CG (i.e., force matching), and then the solvent potential is reintroduced by using an independently derived CG solvent model. This approach yields equilibrium structures that are in better agreement with those produced by the reference atomistic MD simulation than the equilibrium structures generated by the solvent-free CG model. We anticipate that this procedure can be extended to other flexible macromolecules in solution where a MARTINI-like coarse-grained mapping scheme is chemically sensible. Using such an approach may lead to unique physical insight for systems where interactions among macromolecules in solution are important driving forces. Application of this method to investigate how flexible macromolecular conjugates used in drug delivery bind to proteins in solution is currently underway. We caution that this CG approach is not designed to recover dynamic properties, such as transport coefficients. For CG models where accurate time correlations for solutes are desired, the friction and noise forces must be included to the equations of motion, via Langevin dynamics or dissipative particle dynamics. Reference [14] illustrates such a procedure for a CG simulation of a star polymer melt.

Financial support was provided by the National Science Foundation, OCI-0749156 and OCI-0941434. This research was supported in part by Blue Waters, which is a joint effort of the University of Illinois at Urbana-Champaign, its National Center for Supercomputing Applications, IBM, and the Great Lakes Consortium for Petascale Computation. This research also used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the US Department of Energy under contract DE-AC02-06CH11357. We thank Gregory Voth for providing us with the MS-CG code and Lanyuan Lu for assistance with technical issues concerning its implementation. We also thank Dennis Vigil for useful discussions.

[1] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, J. Phys. Chem. B **111**, 7812 (2007).

[2] L. Lu, S. Izvekov, A. Das, H. C. Andersen, and G. A. Voth, J. Chem. Theory Comput. **6**, 954 (2010).

[3] S. H. Kim and M. H. Lamm (unpublished).

[4] S. Izvekov and G. A. Voth, J. Chem. Theory Comput. **2**, 637 (2006).

[5] S. Izvekov and G. A. Voth, J. Phys. Chem. B **113**, 4443 (2009).

[6] Y. Wang and G. A. Voth, J. Phys. Chem. B **114**, 8735 (2010).

[7] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.84.025701 for a direct test of the solvent-free effect and details of the simulation procedure.

[8] R. Baron, A. H. de Vries, P. H. Hunenberger, and W. F. van Gunsteren, J. Phys. Chem. B **110**, 8464 (2006).

[9] W. G. Noid, J. W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008).

[10] W. G. Noid, P. Liu, Y. Wang, J. W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, J. Chem. Phys. **128**, 244115 (2008).

[11] H. Lee and R. G. Larson, J. Phys. Chem. B **112**, 7778 (2008).

[12] Y. Wang, S. Izvekov, T. Yan, and G. A. Voth, J. Phys. Chem. B **110**, 3564 (2006).

[13] M. Lard, S. H. Kim, S. Lin, P. Bhattacharya, P. C. Ke, and M. H. Lamm, Phys. Chem. Chem. Phys. **12**, 9285 (2010).

[14] C. Hijon, P. Espanol, E. Vanden-Eijnden, and R. Delgado Buscalioni, Faraday Discuss. **144**, 301 (2010).