Systematic Coarse Graining of Biomolecular and Soft-Matter Systems

Gary S. Ayton, Will G. Noid, and Gregory A. Voth

Abstract

Coarse-grained modeling is a key component in the field of multiscale simulation. Many biomolecular and otherwise complex systems require the characterization of phenomena over multiple length and time scales in order to fully resolve and understand their behavior. These different scales range from atomic to near macroscopic dimensions, and they are generally not independent of one another, but instead coupled. That is, phenomena occurring at atomic length scales have an effect at macroscopic dimensions and vice versa. Systematic transfer of information between these different scales represents a core challenge in the field of multiscale simulation. Coarse-grained modeling works at an intermediate resolution that can bridge the very high resolution (atomic) scale to the very low resolution (macroscopic) scale. As such, a significant challenge is the development of a systematic methodology whereby coarse-grained models can be derived from their high-resolution atomistic-scale counterpart. Here, a systematic theoretical and computational methodology will be described for developing coarse-grained representations of biomolecular and other soft-matter systems. At the heart of the methodology is a variational statistical mechanical algorithm that uses force-matching of atomistic molecular dynamics data to a coarse-grained representation. A theoretical analysis of the coarse-graining methodology will be presented, along with illustrative applications to membranes, peptides, and carbohydrates.

Introduction

The need for multiscale modeling has been motivated by the inherent design of many complex systems. In order to understand the multiscale nature of a particular complex system, it is necessary to first recast the problem in terms of its multiple spatial and temporal scales. The same system, for example, a protein or a phase-separated fluid, is therefore reconstructed at different levels of spatial resolution. One scale models the system at an atomistic level of resolution, while another scale may model the system as a continuum viscoelastic object. In between, a coarse-grained (CG) resolution is often used, which is part-way between an atomistic and a continuum-level description of the system. The need for coarse-grained representations arises because it is generally quite challenging in complex, inhomogeneous systems to directly bridge the atomistic and continuum-level scales. In between these two disparate regimes, a host of emergent mesoscopic phenomena can occur, and a new intermediate CG representation is required in order to connect the two "end point" scales. As such, coarse graining has become a key component in the full multiscale description of complex systems.

From a computational standpoint, CG modeling allows for significantly larger spatial and temporal scales to be examined, as the CG representation of a system has fewer degrees of freedom than its atomistic counterpart, and the CG interactions are generally more efficient to evaluate on a computer. However, beyond just computational advantages, CG modeling can also provide a powerful tool for understanding the complex multiscale nature of the system. During the process of coarse graining, the description of the system goes from a highly detailed, atomistic-level model to a more reduced representation; if done systematically, important averaged atomistic-level interactions are still retained in the CG model. Then, as the system is examined at lower and lower degrees of resolution, it still retains a critical memory of the original atomistic interactions that governed the system at the highest-resolution scale. It is within this spirit of systematic coarse graining that our approach is cast.

As previously mentioned, the multiscale nature of many biomolecular and soft-matter systems requires modeling strategies that are capable of spanning multiple length and time scales.[1] For example, most biological processes (e.g., membrane remodeling, viral capsid assembly, and ribosomal protein synthesis) occur on length and time scales that approach the macroscale, yet they are tightly coupled to atomic-level phenomena such as protein structure and dynamics. As such, a variety of computational approaches have been developed to examine these systems at various spatial and temporal scales. Molecular dynamics (MD) simulation[2,3] remains arguably the most time-tested tool for investigating the structure and dynamics of biomolecular systems over length and time scales on the order of nanometers and nanoseconds. In order to access the next scale (i.e., micrometers, microseconds), low-resolution, coarse-grained (CG) models can be employed, subject to the degree of actual quantitative information they can produce. In terms of biomolecular systems, for example, CG models have now been employed to examine the mesoscopic behavior of lipid membranes,[4–19] proteins,[20–51] and peptides.[52–55] At the mesoscopic level, for example, emergent phenomena such as long-wavelength thermal undulation modes and domain structures in lipid bilayers are observed. The mesoscopic level occurs at length scales where the system can begin to take on a continuum-level description, yet it is still influenced by strong thermal fluctuations. Various CG models have been previously reviewed for membranes[19] and proteins.[20] A review of the multiscale coupling aspects of these approaches has also been given.[1]

There has been a growing degree of effort to more formally bridge the atomistic

and CG scales.[1] In other words, rather than constructing a "toy" CG model for a complex biomolecular system, current research has been aimed at more systematically constructing CG models directly from atomistic-resolution models or other means. The reverse Monte Carlo method[31] is one such example. The focus of this article, however, will be on our multiscale coarse-graining (MS-CG) methodology,[13–15,53,56–60] which systematically derives highly accurate CG force fields from atomistic MD simulations using a "force-matching" algorithm.[13] The MS-CG methodology connects and bridges atomistic and CG forces via a variational minimization scheme, where CG force fields emerge from the resulting "best fit" as determined from an atomistic-level MD data set. This approach has been employed to model a number of systems[13–15,53,56–58] including pure bilayers,[13] mixed lipid–cholesterol bilayers,[14] peptides,[53] and monosaccharide glucose.[58] This review will highlight some of the key ideas of the underlying methodology as well as provide some key applications. A more detailed account of the underlying theoretical framework can be found elsewhere.[60,61]

Multiscale Coarse-Graining Methodology

The multiscale coarse-graining (MS-CG) method has connections to field-theoretic coarse graining[62] and liquid-state theory.[60,63] Accordingly, some important concepts need to be introduced. In the subsequent analysis, all upper-case variables refer to the CG representation while lower-case variables correspond to the atomistic system. The spatial coordinates of the n atoms of the atomistic system are designated by $\mathbf{r}^n = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, ...\mathbf{r}_n\}$, while the N coarse-grained coordinates are given by $\mathbf{R}^N = \{\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, ...\mathbf{R}_N\}$, with $N < n$. In both cases, the volume of the system is the same and is given by V. The two spatial resolutions are linked via a mapping operator given by

$$\mathbf{M}_{\mathbf{R}}^{N} = \{\mathbf{M}_{\mathbf{R}_1}, \mathbf{M}_{\mathbf{R}_2}, \mathbf{M}_{\mathbf{R}_3}, ..., \mathbf{M}_{\mathbf{R}_N}\}. \tag{1}$$

In the case that a center of mass transformation from the atomistic to CG coordinates is desired, then

$$\mathbf{M}_{\mathbf{R}_I}(\mathbf{r}^n) = \frac{\sum_{i=1}^{n} P_{Ii} m_i \mathbf{r}_i}{\sum_{i=1}^{n} P_{Ii} m_i}, \tag{2a}$$

where $P_{Ii} = 1$ if atom i is part of the CG site I and is zero otherwise. The subscripts on P denote which atom is associated with a particular CG site. Here, the mass of atom i is m_i, while the mass of the CG site is

$$M_I = \sum_{i=1}^{n} P_{Ii} m_i. \tag{2b}$$

With this transformation from atomistic to CG coordinates, it can be defined that there exists a multidimensional CG potential of mean force (PMF), $U(\mathbf{R}^N)$, that is rigorously related to the potential energy function of the atoms $u(\mathbf{r}^n)$ via

$$U(\mathbf{R}^N) = -k_B T \ln \left[\frac{Z_N}{z_n} \int d\mathbf{r}^n \prod_{I=1}^{N} \delta\left(\mathbf{R}_I - \mathbf{M}_{\mathbf{R}_I}(\mathbf{r}^n)\right) e^{-\beta u(\mathbf{r}^n)} \right], \tag{3}$$

where k_B is the Boltzmann constant, T is the thermodynamic temperature, $\beta = 1/k_B T$, Z_N is the CG configurational integral, z_n is the atomistic configurational integral, and the delta functions δ are only nonzero when $\mathbf{R}_I = \mathbf{M}_{\mathbf{R}_I}(\mathbf{r}^n)$. A corresponding canonical transformation can also be employed for Equation 3.[60] The two configurational integrals are, respectively,

$$Z_N = \int d\mathbf{R}^N e^{-\beta U(\mathbf{R}^N)} \tag{4a}$$

and

$$z_n = \int d\mathbf{r}^n e^{-\beta u(\mathbf{r}^n)}. \tag{4b}$$

The form of $U(\mathbf{R}^N)$ as given in Equation 3 arises from the initial requirement that the excess free energy of the CG system, A_{ex}, is the same as that for the atomistic system, a_{ex}; that is, $a_{ex} = A_{ex}$. The CG simulation will also produce the correct spatial distribution of CG sites. In both cases, the excess free energies for the atomistic and CG system are given by, respectively,

$$a_{ex} = -k_B T \ln\left[\frac{z_n}{V^n}\right] \tag{5a}$$

and

$$A_{ex} = -k_B T \ln\left[\frac{Z_N}{V^N}\right]. \tag{5b}$$

The CG many-dimensional PMF $U(\mathbf{R}^N)$ cannot be easily evaluated from computer simulations of the atomic system. This can be demonstrated by considering a specific CG configuration $\mathbf{R}^N$, where this configuration gives a particular spatial arrangement of the CG sites. Now imagine performing an atomistic-level MD simulation where the CG sites are related to the atomistic coordinates via Equation 2. Each time the atomic coordinates map onto the CG coordinates, the product of delta functions in Equation 3 is nonzero, and the left-hand side of Equation 3 results in a single numerical value (this is called a "hit"). Each time the atomic configuration maps onto the preselected CG configuration, another hit occurs and another number is collected. Over many sampled atomistic configurations, the *numerical* value of $U(\mathbf{R}^N)$ for that specific $\mathbf{R}^N$ is found. However, no other information—that is, the functional form of $U(\mathbf{R}^N)$—is provided. As such, from first principles, this approach to constructing efficient CG models would very quickly prove intractable.

However, it is possible to work with the gradients of $U(\mathbf{R}^N)$; this approach is employed, then A_{ex} will be related to a_{ex} to within a constant that is independent of the spatial coordinates of either the atomistic or CG system. The forces that act on the CG coordinates are found from gradients of $U(\mathbf{R}^N)$ as

$$\mathbf{F}_I(\mathbf{R}^N) \equiv -\nabla_{\mathbf{R}_I} U(\mathbf{R}^N), \tag{6}$$

where $\mathbf{F}_I$ represents a force; it can be shown that these forces can be calculated as specially selected ensemble averages of atomistic forces in an equilibrium MD simulation.[13,60,61] For example, in most cases, $\mathbf{F}_I$ is found from the total force arising from the set of atoms involved in the CG site at $\mathbf{R}_I$, averaged over all the atomistic configurations where the CG coordinates map onto $\mathbf{R}^N$ via Equation 2. Mathematically, this is expressed as

$$\mathbf{F}_I(\mathbf{R}^N) = \langle \mathbf{f}_I(\mathbf{r}^n) \rangle_{\mathbf{R}^N}, \tag{7a}$$

where

$$\mathbf{f}_I = \sum_{i=1}^{n} P_{I,i} \mathbf{f}_i, \tag{7b}$$

$\mathbf{f}_i$ is the total atomistic force on atom i, and the $\mathbf{R}^N$ subscript on the average denotes that the average is evaluated only over those atomistic configurations that map onto $\mathbf{R}^N$ via Equation 2.

The MS-CG method approximates $\mathbf{F}_I$ with an effective force, $\mathbf{F}_I^{MS}$. In the case of nonbonded interactions, for example, the effective force can have the form of a pairwise radial interaction over the surrounding CG sites, such that

$$\mathbf{F}_I^{MS} = \sum_{J \neq I}^{N} \mathbf{F}_{IJ}^{MS}, \tag{8}$$

where $\mathbf{F}_{IJ}^{MS}$ is an effective pair force between CG site I and CG site J. The effective pair force is linearly dependent on N_k parameters[13] denoted by the vector

$$\mathbf{l} = \{l_1, l_2, l_3 \ldots l_{N_k}\}. \tag{9}$$

The set of N_k parameters model the distance-dependence of $\mathbf{F}_{IJ}^{MS}$ and can take the form of, for example, spline coefficients.[13] The optimal $\mathbf{F}_{IJ}^{MS}$ is found from force-matching,[13] which relies on the variational minimization of a residual function, χ^2_{MS}, defined as

$$\chi^2_{MS}(\mathbf{l}) = \frac{1}{3n_t N} \sum_{t=1}^{n_t} \sum_{I=1}^{N} \left| \mathbf{f}_I(\mathbf{r}_t^n) - \mathbf{F}_I^{MS}\left(\mathbf{M}_{\mathbf{R}}^N(\mathbf{r}_t^n); \mathbf{l}\right) \right|^2, \tag{10}$$

where n_t is the total number of configurations sampled, and $\mathbf{r}_i^n$ corresponds to the tth atomically detailed configuration. The best solution is found at the minimum of the residual, that is, where $\nabla_I \chi^2_{MS} = 0$. The fact that the residual is quadratic in its parameters guarantees that only one minimum can exist. The result is a pairwise additive tabulated radial force field that can be used in a subsequent MS-CG simulation. More complicated systems with many components and bonded interactions can also be considered (e.g., a lipid bilayer or peptide).[13,14,53]

Furthermore, it can be shown that the set of normal linear equations found from evoking the $\nabla_I \chi^2_{MS} = 0$ maps over to the Yvon–Born–Green (YBG) equation[60,63] in certain isotropic systems and thus incorporates information from three-body correlations, which are critical in order to accurately define a CG force field.

The MS-CG simulation often employs CG "Hamiltonian" dynamics where, in the simplest case,

$$\dot{\mathbf{R}}_I = \mathbf{P}_I / M_I \tag{11a}$$

and

$$\dot{\mathbf{P}}_I = \mathbf{F}_I^{MS}, \tag{11b}$$

with $\mathbf{P}_I$ being the appropriate CG momenta, that is, the mass of the CG site multiplied by the velocity of the site. These equations of motion can be further augmented to maintain, for example, isothermal conditions using methodologies employed in typical atomistic MD simulations.[64] Importantly, it is also possible to incorporate the random and frictional forces that are required to obtain a more realistic CG dynamics.[59]

Representative Applications

In the following subsections, some representative applications of the MS-CG method are presented, including results for lipid bilayers, peptides, and carbohydrates (i.e., a monosaccharide glucose solution). The initial stages of this work have focused on validating the MS-CG approach directly against atomistic simulation. However, we are now employing MS-CG to examine larger length- and time-scale phenomena like aggregation with quantitative accuracy. Future work will incorporate the MS-CG method into a fully multiscale simulation framework, bridging the atomistic to continuum (near-macroscopic) scales. Algorithmic details of the MS-CG method can be found elsewhere,[13,14,53,58] and here the focus will be on the results. In Figure 1, the various CG schemes for the different systems are shown.

Lipid Bilayers

The MS-CG method was first utilized to obtain a CG model of a dimyristoylphosphatidylcholine (DMPC) lipid bilayer.[13] At the atomistic level, a single atomistic-level MD simulation of a fully solvated lipid bilayer was employed in order to determine the required MS-CG force fields resulting from the CG decomposition of the lipid, as depicted in Figure 1a. Details of the atomistic simulations and the atomistic force fields can be found elsewhere.[65] A subsequent study employed a mixed lipid-cholesterol CG model at 50 mol% cholesterol concentration,[14] where two different CG representations of the cholesterol molecule were employed, as shown in Figure 1b.

The MS-CG method results in force fields between different CG sites. In turn, when the system is reconstructed using the CG sites instead of the original atoms, the new CG forces acting on these sites result in CG structural correlations. If done systematically and accurately, the structural correlations between the CG sites should be the same as those observed in the original atomistic simulation. Structural correlations can be quantitatively measured via radial distribution functions (RDFs), which give the probability of observing a CG site of type α at a given distance r, given that there is a CG site of type β at the origin. As such, comparing RDFs from the MS-CG and original MD simulation acts as a quantitative measure of how well the CG model is describing the original atomistic-level system at the new CG resolution. Some representative MS-CG RDFs for the seven-site cholesterol CG model are given in Figure 2, where the resulting RDFs are in good agreement with the atomistic-level MD simulation results. This is a critical test: the MS-CG force field, when employed in a subsequent CG simulation, should preserve the structure that was observed in the all-atom MD simulation. Small deviations in the structure can result from variations in the three-body correlations in the all-atom simulation, as compared to the MS-CG simulation.[60] They can also arise due to the assumption of a pairwise additive radial CG force field in the MS-CG method. This particular application was also quite encouraging in terms of the MS-CG method, because good CG models for cholesterol have been

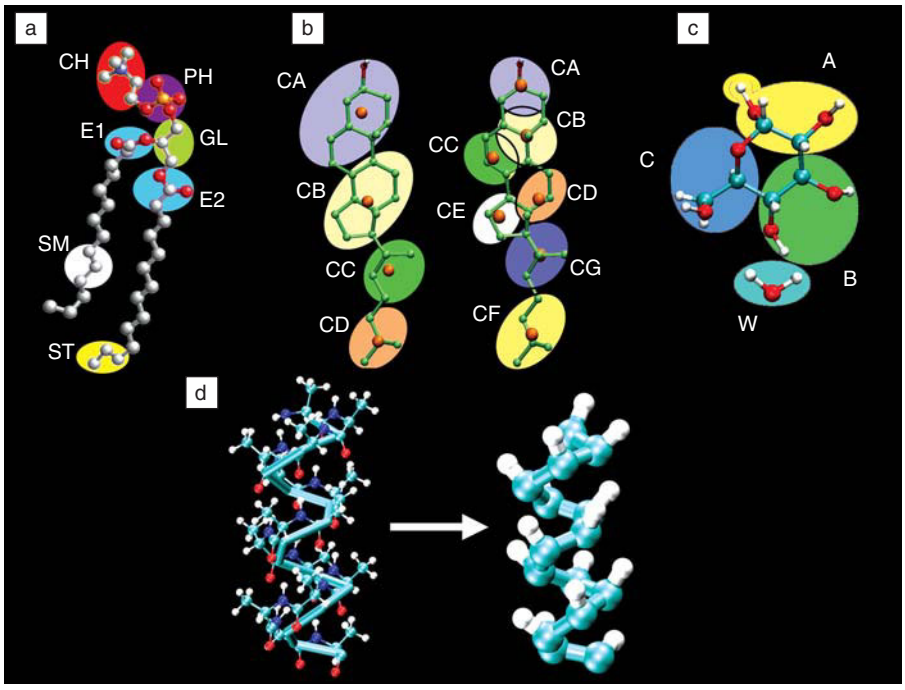

Figure 1. Various coarse-graining schemes for (a) lipids; (b) cholesterol, 4-site and 7-site models; (c) carbohydrates; and (d) peptides (helical polyalanine). The various CG site designations are shown in the figure (i.e., CH, PH, GL, etc.). These labels designate CG sites for a particular coarse-graining scheme, and do not necessarily correspond to distinct molecular groups. The peptide is shown in full atomistic detail on the left image (d), while in the CG model on the right, two sites are employed for each residue: one for the peptide backbone (green) and another for the side-chain (white). "W" stands for CG water.

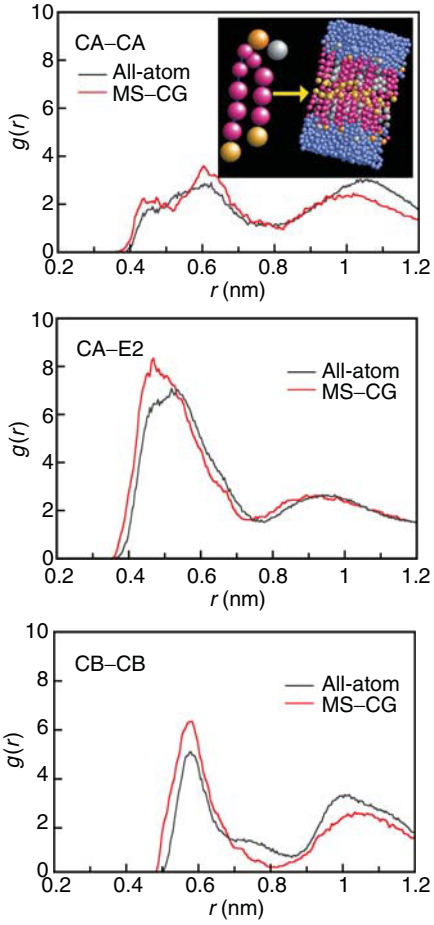

Figure 2. Selected radial distribution functions $g(r)$, versus intersite CG distance r, for a lipid–cholesterol mixed-bilayer system, where the black curves are the all-atom molecular dynamics simulation results and the red curves are the multiscale coarse-graining (MS-CG) pair distribution functions (for the 7-site CG model). CG site labels are illustrated in Figure 1. The inset in the top panel shows a model of the pure CG bilayer with the surrounding CG solvent.

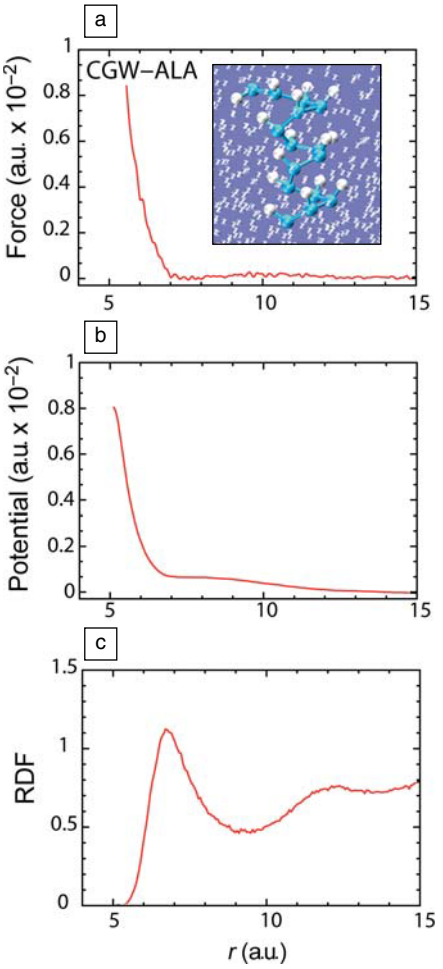

Figure 3. (a) Force, (b) potential, and (c) radial distribution function curves versus radial distance r for the CGW–ALA peptide interaction. The inset in the top image shows the coarse-grained representation of a 15-residue polyalanine helix in explicit CG water. The peptide backbone and side-chain groups are displayed in cyan and white, respectively. The CG solvent molecules are shown as translucent spheres.

very difficult to develop using other methods. The resulting solvated CG lipid bilayer is shown in the inset of Figure 2.

Peptides

The MS-CG methodology has also been employed to develop CG force fields for peptides,[53] employing a polyalanine pentadecamer as the underlying MD model. Specific details of the simulations can be found elsewhere.[53] The CG scheme is shown in Figure 1d. The CG water–alanine (CGW–ALA) MS-CG results are shown in Figure 3, where the correlations between the MS-CG force field, the resulting potential, and the pair distribution functions are evident. Interestingly, the resulting MS-CG potential does not exhibit the typical repulsive–attractive behavior of most atomistic force fields. The resulting RDFs follow from this behavior. Future research will focus on utilizing these MS-CG force fields to diagnose features that can define peptide and protein conformation.

Importantly, overall it was observed from the simulations that the MS-CG models gave a backbone rms deviation for the native state of polyalanine to within ~1 Å, indicating that the MS-CG force field preserved the critical stability of the helix. A snapshot of the system, including the CG water, is shown in the inset of Figure 3a.

Carbohydrates

The MS-CG methodology has been applied to a 9.1 mol% monosaccharide glucose solution;[58] the CG representation is given in Figure 1c. For this study, a number of thermodynamic quantities were also calculated at both the atomistic and the CG level. The isothermal compressibility κ_T defined from

$$\kappa_T = -(1/V)(\partial V/\partial P)_T, \quad (12)$$

where V is the volume and P is the pressure, was evaluated at T = 300 K for the MS-CG model and was determined to be 4.7×10^{-5} atm^{-1}. The CG value is larger than that found from the corresponding atomistic MD case, where $\kappa_T \cong 1.5 \times 10^{-5}$ atm^{-1}. The discrepancy can largely be traced back to the fact that the pressure in a CG simulation is, strictly speaking, not as easy to calculate as in atomistic simulations. Current research along these lines is underway. The thermal expansion coefficient α can be evaluated from

$$\alpha = (1/V)(\partial V/\partial T)_p, \quad (13)$$

where T is the temperature. In contrast to κ_T, α from the CG simulations was found to be 5.6×10^{-4} K^{-1}, in good agreement with that found from the atomistic simulations at 6.5×10^{-4} K^{-1}. This result suggests that this MS-CG model may exhibit reasonable transferability at different temperatures. The self-diffusion coefficient was found to be an order of magnitude larger compared with that from the atomistic MD simulations. This is not surprising, and is in fact expected when the smooth MS-CG force field is used under CG Hamiltonian equations of motion. More realistic CG dynamics can be recovered, however, when the appropriate friction forces are included in the CG dynamics.[59] However, in situations where only structural information is of interest, the inherent accelerated dynamics associated with using the CG Hamiltonian form can, in fact, be advantageous.

Selected force, energy, and RDFs for the A–W and A–C CG site interactions in Figure 1c are shown in Figure 4. (A "core" region was added in this study to ensure that the CG sites could not penetrate into each other.) The A–W interactions, in particular, exhibit interesting details that ultimately emerge in the CG RDF. In general, quantitative agreement between the atomistic and CG representations is observed. A model of the CG system is shown in Figure 5, where the both the CG solvent and CG

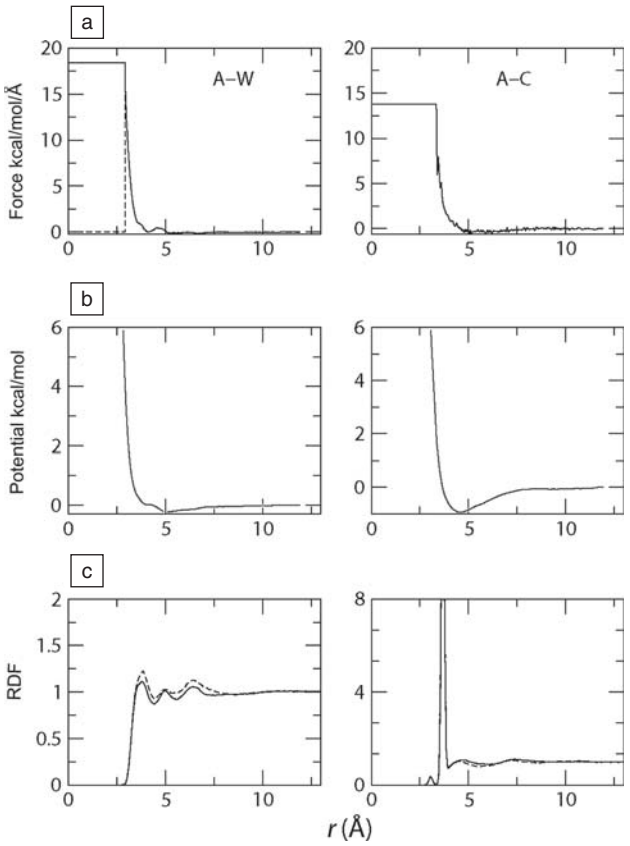

Figure 4. (a) Representative effective force, (b) energy curves, and (c) radial distribution functions as a function of site pair distance *r*. CG A–W (left column) and A–C (right column) sites, employed in the multiscale coarse-graining simulations of a carbohydrate (see Figure 1) are given. In the top-left subplot, the dashed line is the force curve directly from the force-matching method, while the solid line is the force curve with a constant extrapolation into the core region.

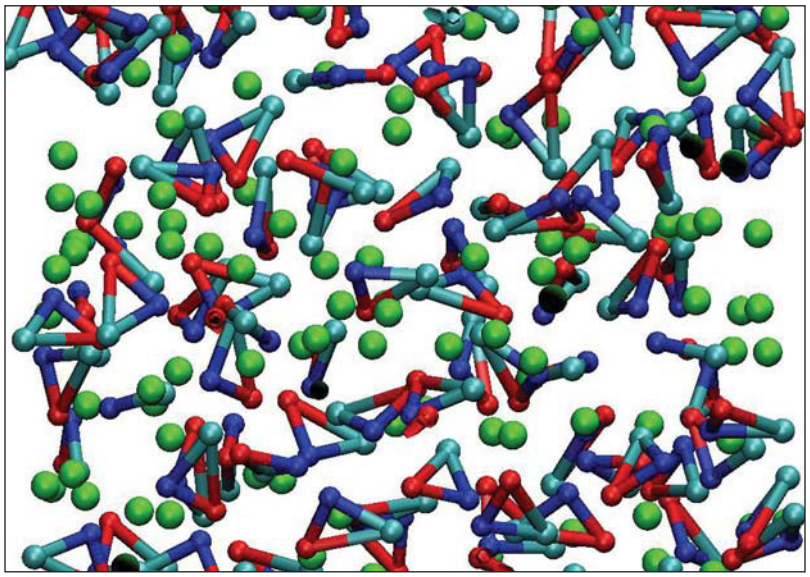

Figure 5. Model of a CG alpha-D-glucopyranose aqueous solution. The green, cyan, blue, and red spheres correspond to the CG water site and the CG A, B, and C carbohydrate sites, respectively.

alpha-D-glucopyranose are shown. The general complexity and size of this system would be extremely difficult, if not impossible, to study with present-day atomistic molecular dynamics methods, yet quantitative accuracy is still maintained in the MS-CG model, as seen in Figure 4.

Summary

The multiscale coarse-graining methodology is a robust coarse-graining approach that employs atomistic-level information to systematically construct CG force fields that can be used in subsequent CG simulations. The methodology is based on a rigorous theoretical framework that can directly bridge atomistic-level interactions over to a lower-resolution CG potential. Importantly, MS-CG exploits some very important, and perhaps not immediately apparent, aspects of the problem that enable the calculation of CG force fields without ever having to resolve the many-dimensional CG potential of mean force.

A number of representative applications were presented in order to demonstrate the flexibility and accuracy of the MS-CG methodology. From the onset, the methodology was developed with the goal of describing complex biomolecular and soft-matter systems. As such, the soft-matter character, as well as the electrostatic interactions that typically dominate these systems, has been incorporated into the methodology from the onset.

Future work will extend the theoretical formalism underlying the MS-CG method, as well as apply it to even more complex systems at various CG resolutions.

Acknowledgments

This research was supported by the National Science Foundation (CHE-0628257) and the National Institutes of Health (5-R01-GM063796). We thank Drs. Sergei Izvekov, Jhih-Wei Chu, Pu Liu, Yanting Wang, Ian Thorpe, and Jian Zhou for valuable discussions and contributions. W.G. Noid acknowledges the support of the National Institutes of Health through a Ruth L. Kirschstein NRSA postdoctoral fellowship grant (5F32GM076839–02).

References

1. G.S. Ayton, W.G. Noid, G.A. Voth, *Curr. Opin. Struct. Bio.* **17**, 192 (2007).
2. M. Karplus, J.A. McCammon, *Nat. Struct. Bio.* **9**, 646 (2002).
3. S.E. Feller, *Curr. Opin. Colloid Interface Sci.* **5**, 217 (2000).
4. R. Goetz, R. Lipowsky, *J. Chem. Phys.* **108**, 7397 (1998).
5. M.J. Stevens, *J. Chem. Phys.* **121**, 11942 (2004).
6. J.C. Shelley, M.Y. Shelley, R.C. Reeder, S. Bandyopadhyay, M.L. Klein, *J. Phys. Chem. B* **105**, 4464 (2001).

7. J.C. Shelley, M.Y. Shelley, R.C. Reeder, S. Bandyopadhyay, M.L. Klein, *J. Phys. Chem. B*. **105**, 9785 (2001).
8. S.J. Marrink, A.H. deVries, A.E. Mark, *J. Phys. Chem. B* **108**, 750 (2004).
9. S.J. Marrink, A.E. Mark, *J. Amer. Chem. Soc*. **125**, 11144 (2003).
10. A.Y. Shih, A. Arkhipov, P.L. Freddolino, K. Schulten, *J. Phys. Chem. B* **110**, 3674 (2006).
11. I.R. Cooke, K. Kremer, M. Deserno, *Phys. Rev. E* **72**, 011506 (2005).
12. O. Farago, *J. Chem. Phys*. **119**, 596 (2003).
13. S. Izvekov, G.A. Voth, *J. Phys. Chem. B* **109**, 2469 (2005).
14. S. Izvekov, G.A. Voth, *J. Chem. Theor. Comp*. **2**, 637 (2006).
15. Q. Shi, S. Izvekov, G.A. Voth, *J. Phys. Chem. B* **110**, 15045 (2006).
16. G.S. Ayton, G.A. Voth, *J. Struct. Bio*. **157**, 570 (2007).
17. A.P. Lyubartsev, *Eur. J. Biophys*. **35**, 53 (2005).
18. G. Brannigan, F.L.H. Brown, *J. Chem. Phys*. **120**, 1059 (2004).
19. G. Brannigan, L.C.L. Lin, F.L.H. Brown, *Eur. Biophys. J*. **35**, 104 (2006).
20. V. Tozzini, *Curr. Opin. Struc. Bio*. **15**, 144 (2005).
21. M. Levitt, A. Warshel, *Nature* **253**, 694 (1975).
22. M. Levitt, *J. Mol. Bio*. **104**, 59 (1976).
23. S. Tanaka, H.A. Scheraga, *Macromolecules* **9**, 945 (1976).
24. S. Miyazawa, R.L. Jernigan, *Macromolecules* **18**, 534 (1985).
25. M.J. Sippl, *J. Mol. Bio*. **213**, 859 (1990).
26. N.V. Buchete, J.E. Straub, *J. Chem. Phys*. **118**, 7658 (2003).
27. M. Nanias, M. Chinchio, S. Oldziej, C. Czaplewski, H.A. Scheraga, *J. Comput. Chem*. **26**, 1472 (2005).
28. M. Nanias, C. Czaplewski, H.A. Scheraga, *J. Chem. Theory Comput*. **2**, 513 (2006).
29. R.L. Jernigan, I. Bahar, *Curr. Opin. Struct. Bio*. **6**, 195 (1996).
30. M. Vendruscolo, E. Domany, *J. Chem. Phys*. **109**, 11101 (1998).
31. S. Matysiak, C. Clementi, *J. Mol. Bio*. **343**, 235 (2004).
32. J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, *Protein. Struct. Funct. Genet*. **21**, 167 (1995).
33. N. Go, *Annu. Rev. Biophys. Bioeng*. **12**, 183 (1983).
34. E. Kussel, J. Shimada, E.I. Shakhnovich, *Proc. Natl. Acad. Sci. USA* **99**, 5343 (2002).
35. Y. Fujitsuka, S. Takada, Z.A. Luthey-Schulten, P.G. Wolynes, *Struct. Funct. Bioinform*. **54**, 88 (2004).
36. F. Ding, W. Guo, N.V. Dokholyan, E.I. Shakhnovich, J.E. Shea, *J. Mol. Biol*. **350**, 1035 (2005).
37. M.M. Tirion, *Phys. Rev. Lett*. **77**, 1905 (1996).
38. T. Haliloglu, I. Bahar, B. Erman, *Phys. Rev. Lett*. **79**, 3090 (1997).
39. A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, I. Bahar, *Biophys. J*. **80**, 505 (2001).
40. T.Z. Sen, Y. Feng, J.V. Garcia, A. Kloczkowski, R.L. Jernigan, *J. Chem. Theory Comput*. **2**, 696 (2006).
41. O. Kurkcuoglu, R.L. Jernigan, P. Doruker, *Biochemistry* **45**, 1173 (2006).
42. Z. Zhang, Y. Shi, H. Liu, *Biophys. J*. **84**, 3583 (2003).
43. N. Go, H. Taketomi, *Proc. Natl. Acad. Sci. USA* **75**, 559 (1978).
44. J. Shimada, E.L. Shakhnovich, *Proc. Natl. Acad. Sci. USA* **99**, 11175 (2002).
45. S. Brown, N.J. Fawzi, T. Head-Gordon, *Proc. Natl. Acad. Sci. USA* **100**, 10712 (2003).
46. D.M. Zuckerman, *J. Phys. Chem. B* **108**, 5127 (2004).
47. D. Swigon, B.D. Coleman, W.K. Olson, *Proc. Natl. Acad. Sci. USA* **103**, 9879 (2006).
48. M. Lu, B. Poon, J. Ma, *J. Chem. Theory Comput*. **2**, 464 (2006).
49. S. Matysiak, C. Clementi, *J. Mol. Biol*. **363**, 297 (2006).
50. J.-W. Chu, G.A. Voth, *Biophys. J*. **90**, 1572 (2006).
51. A. Liwo, C. Czaplewski, J. Pillardy, H.A. Scheraga, *J. Chem. Phys*. **115**, 2323 (2001).
52. S. Peng, F. Ding, B. Urbanc, S.V. Buldyrev, L. Cruz, H.E. Stanley, N.V. Dokholyan, *Phys. Rev. E*. **69**, 041908 (2004).
53. J. Zhou, I.F. Thorpe, S. Izvekov, G.A. Voth, *Biophys. J*. **92**, 4289 (2007).
54. A.E. van Giessen, J.E. Straub, *J. Chem. Theory Comput*. **2**, 674 (2006).
55. A.E. van Giessen, J.E. Straub, *J. Chem. Phys*. **122**, 024904 (2005).
56. S. Izvekov, G.A. Voth, *J. Chem. Phys*. **123**, 134105 (2005).
57. Y. Wang, S. Izvekov, T. Yan, G.A. Voth, *J. Phys. Chem. B* **110**, 3564 (2006).
58. P. Liu, S. Izvekov, G.A. Voth, *J. Phys. Chem. B* (2007) in press.
59. S. Izvekov, G.A. Voth, *J. Chem. Phys*. **125**, 151101 (2006).
60. W.G. Noid, J.-W. Chu, G.S. Ayton, G.A. Voth, *J. Phys. Chem. B* **111**, 4116 (2007).
61. J.-W. Chu, G.S. Ayton, S. Izvekov, G.A. Voth, *Mol. Phys*. **105**, 167 (2007).
62. P.M. Chaikin, T.C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, U.K., 1995).
63. J.P. Hansen, I.R. McDonald, *Theory of Simple Liquids* (Academic Press, San Diego, ed. 2, 1986).
64. D.J. Evans, B.L. Holian, *J. Chem. Phys*. **83**, 4069 (1985).
65. A.M. Smondyrev, M.L. Berkowitz, *J. Comp. Chem*. **20**, 531 (1999). □