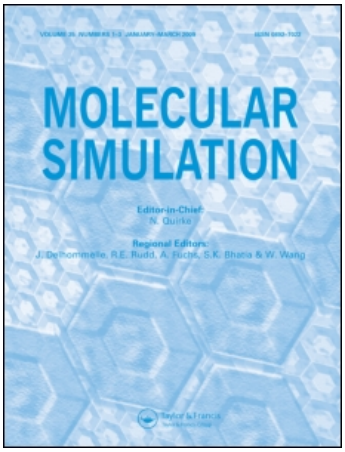

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

The multiscale challenge for biomolecular systems: coarse-grained modeling

J. -W. Chu[a]; S. Izveko[a]; G. A. Voth[a]
[a] Department of Chemistry, Center for Biophysical Modeling and Simulation, University of Utah, Salt Lake City, UT, USA

To cite this Article Chu, J. -W. , Izveko, S. and Voth, G. A.(2006) 'The multiscale challenge for biomolecular systems: coarse-grained modeling', Molecular Simulation, 32: 3, 211 — 218
To link to this Article: DOI: 10.1080/08927020600612221
URL: http://dx.doi.org/10.1080/08927020600612221

Molecular Simulation, Vol. 32, Nos. 3–4, 15 March–April 2006, 211–218

Taylor & Francis
Taylor & Francis Group

The multiscale challenge for biomolecular systems: coarse-grained modeling

J.-W. CHU, S. IZVEKO and G. A. VOTH*

Department of Chemistry, Center for Biophysical Modeling and Simulation, University of Utah, 315 S. 1400 E. Rm 2020, Salt Lake City, UT 84112-0850, USA

(Received January 2006; in final form February 2006)

Two new approaches are presented for obtaining coarse-grained (CG) force fields from atomistic molecular dynamics (MD) trajectories. The first approach is the force-matching (FM) method whereby the force data obtained from an explicit atomistic MD simulation are utilized to determine the CG force fields. The performance of the FM method is demonstrated by applying it to derive a CG model for the dimyristoylphosphatidylcholine (DMPC) lipid bilayer. The second approach is the fluctuation-matching method whereby the fluctuations of specific internal coordinates are extracted from atomistic MD simulations to derive the CG force field. The fluctuation matching method is then applied to analyze the mechanical behavior of actin filaments. Both methods propagate the information obtained at a fine-grained atomistic scale to a CG scale, and hence are termed as multiscale coarse-graining approaches. Such multiscale approaches provide a new route to model complex biomolecular systems.

Keywords: Multiscale; Coarse-grain; Force matching; Fluctuation matching; Biological membrane; Lipid bilayer

1. Introduction

Complex biological assemblies such as lipid bilayers, the DNA double helix, and the cytoskeleton are the building blocks of living systems. The multiple and disparate time- and length-scales that occur in these assemblies are the major challenges in developing a molecular-level understanding of cell biology. The coupling between microscopic interactions among atoms and macroscopic and mesoscopic properties, such as bending modulus, undulation spectrum, and raft formation, require simulation methods capable of bridging information at different spatial and temporal regimes. This multiscale challenge has motivated the development of so-called "coarse-grained" (CG) approaches [1–5]. The philosophy of CG approaches is generally the same: to achieve a simpler description of the effective interactions in a given system while enhances the ability of the resulting models to predict the properties of interest. After the details in a fine-grained atomistic description are reduced, the accessible time- and length- scales of molecular simulations can be increased. If the physics of the interested processes are captured, CG simulations could be used to systematically study complicated biological systems.

A typical way of achieving a simpler description through coarse-graining is to reduce the structural details of a complicated system by grouping atoms into fewer interaction sites (CG sites) as in figure 1 for lipid molecules. Therefore, the development of CG models usually involves two distinct steps: the first step is to group the degrees of freedom in a system into fewer structural units of CG sites. The second step is the construction of an effective force field to describe the interactions among these CG sites. The first step generally relies on the chemical intuition of modelers, while the second stage represents a major difficulty in CG modeling because it is very difficult to know a priori all interactions of a system that need to be included in a CG model. Therefore, it is important to be able to systematically develop a CG model based on the underlying interactions and details from a fine-grained (atomistic-scale) description. However, there have generally been few systematic strategies for the development of CG potentials. Typically, CG potentials of a pre-selected analytical form are parameterized either to reproduce certain structural properties by using, for example, an iterative adjustment of potential parameters or the inverse Monte Carlo technique, or they are parameterized by matching with the

*Corresponding author. Tel: 1 801 581 7272. Fax: 1 801 581 4353. Email: voth@chem.utah.edu

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020600612221

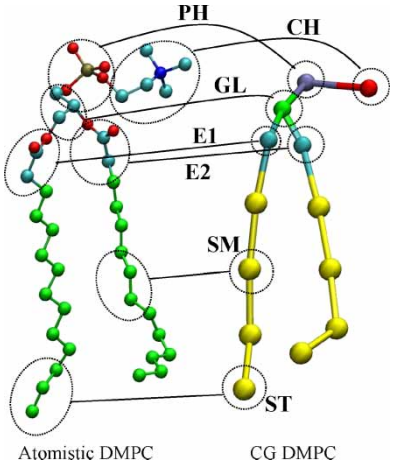

Figure 1. Atomistic (left) and CG (right) representations of a DMPC molecule. The CG units are the centers-of-mass of: choline (CH), phosphate (PH), glycerol backbone (GL: $CH_2-CH-CH_2$), ester groups (E1 and E2: O_2CCH_2), the triplets of carbon atoms of alkane chains [SM: $(CH_2)_3$], and the tail of alkane chains [ST: $(CH_2)_2-CH_3$].

desired thermodynamic properties [6–8]. These approaches are not directly based on the underlying atomistic scale interactions. Another issue of CG modeling is the transferability of force fields; the parameters that are obtained from the parameterization under certain thermodynamic or structural conditions may not be applicable to others. As a result, re-parameterization of CG force fields may become a common need and practice in the multiscale modeling and simulation of bioassemblies.

The aforementioned challenges have motivated our development of systematic procedures for CG simulations of biological systems [9,10]. The goal is to systematically extract CG parameters directly from the underlying fine-grained interactions (usually all-atom models with empirical force fields). In this way, the information at the fine-grained level can be propagated to the CG level and CG parameters could be obtained specifically for the conditions that are of interest. We denote such approaches as multiscale coarse-graining (MS-CG) approaches. In this article, two MS-CG approaches, the force-matching (FM) and fluctuation-matching methods, are introduced with applications to lipid bilayers and actin filaments.

In the following section, the essence of the FM and fluctuation-matching methods is introduced followed by the results of applying these methods to study the biophysical properties of lipid bilayers and actin filaments. Finally, concluding remarks will be drawn.

2. Methods

2.1 Force-matching

The forces on CG sites are determined by the force field parameters that describe the interactions among these sites. The force field parameters are denoted as $(g_1, \ldots, g_M)$; M is the total number of force field parameters in a CG model, and is determined by the modeler. The forces on a CG site I can thus be determined by the force field parameters: $(g_1, \ldots, g_M)$. In the FM method, the set of M parameters, $(g_1, \ldots, g_M)$, are determined by minimizing the average error in forces over the whole configuration data in an atomistic molecular dynamics (MD) simulation:

$$\frac{1}{3LN}\sum_{l}\sum_{I}\left|F_{I,l}^{\mathrm{ref}} - F_{I,l}^{CG}\right|^2. \tag{1}$$

In equation (1), $F_{I,l}^{\mathrm{ref}}$ is the force on CG site I computed from atomistic force fields at configuration l in an MD trajectory, and $F_{I,l}^{CG}$ is the force computed from CG force field parameters at the same configuration [11]. Therefore, for an CG representation of N sites and an atomistic MD simulation of L configurations, the averaged error in forces caused by using CG force fields compared to the forces computed from atomistic force fields is defined as in equation (1). Notice that the dependence of CG forces on force field parameters; $F_I^{CG}(g_1, \ldots, g_M)$, can be specified arbitrarily by the user. However, the fitting rapidly becomes intractable as the number of parameters grows. For biological molecules, for example, the problem becomes extreme because of the large variety of atomic species that could be found in a system. However, if the force field to be fit depends linearly on the fitting parameters, such as those that can be often achieved using a suitable (e.g. spline) interpolation method, the least squares problem can be written in the form of an overdetermined system of linear equations. The least-squares solution of such system can then be efficiently found using, for example, orthogonal matrix triangularization (QR decomposition) or singular value decomposition [11].

In the molecular modeling of soft matter, the interactions among CG sites are usually assumed to be pair wise decomposable

$$F_I^{CG} = \sum_{J\neq I} f_{IJ}^{CG}\left(\left|\vec{R}_{IJ}\right|\right)\frac{\vec{R}_{IJ}}{\left|\vec{R}_{IJ}\right|}, \tag{2}$$

where $\vec{R}_{IJ}$ in the displacement vector from site I to J. $f_{IJ}^{CG}\left(\left|\vec{R}_{IJ}\right|\right)$ is the force on site I due to site J, and is a function only depends on the distance between sites I and J, $\left|\vec{R}_{IJ}\right|$. In order to exploit the numerical efficiency of solving linearized least-square problems, $f_{IJ}^{CG}\left(\left|\vec{R}_{IJ}\right|\right)$ is represented by using a numerical table at a function of $\left|\vec{R}_{IJ}\right|$ with spline interpolation to ensue the continuity of the derivatives of $f_{IJ}^{CG}\left(\left|\vec{R}_{IJ}\right|\right)$. The tabulated form of forces allows one to carry out the fit in a manner that is more consistent with the determination of the mean CG force field (site–site potential of mean force or PMF). In this case, the least square problem is solved for small blocks of atomic trajectories and the solutions obtained are averaged over a large number of such blocks. Typically, the number of atomic configurations used to build the least square problem in each block is sought to ensure that these

equations overdetermine the force field parameters and the least square problem in each block has a unique solution. More details about the force matching method can be found in Izvekov *et al.* [11].

2.2 Fluctuation-matching

The FM method is a general way of obtaining CG force fields directly from atomistic simulations, and the use of numerical tables eliminate the restriction imposed by assuming certain functional forms for the CG potential. Therefore, the FM method is an optimal choice when the nature of the underlying interactions among CG sites is complicated and when there are not too many such sites.

However, in many cases it is desirable to adopt certain forms of interactions among CG sites so to relate them directly to the material properties of a bioassembly. For example, by using effective harmonic interactions among CG sites, one can relate to the force constants to the elastic modulus of a lipid bilayer [12]. Even under the circumstance of assuming particular forms of interactions, it is still desirable to be able to bridge the information obtained from fine-grained simulations to the mesoscopic CG model. The fluctuation-matching method is thus a way to systematically extract the information of CG parameters from atomistic simulations by employing the concept of quasiharmonic analysis [13]. Fluctuation-matching is particularly efficient, but not limited to, when the interactions among CG sites are assumed to be effective harmonic interactions [10].

For each effective internal coordinate, EIC, that the modeler defines to describe the interactions among CG sites, the equilibrium value ($\langle \mathrm{EIC} \rangle$) and fluctuation around the mean:

$$\left\langle \delta \mathrm{EIC}_i^2 \right\rangle = \left\langle \left(\mathrm{EIC}_i - \langle \mathrm{EIC}_i \rangle \right)^2 \right\rangle \tag{3}$$

in a fine-grained MD simulation can be obtained by analyzing the trajectory. For each EIC, the fluctuation can also be computed for the CG potential via MD simulations or normal mode analysis (NMA), and the effective force fields in a CG model can thus be determined by matching the fluctuations computed from the CG model to that obtained from explicit fine-grained data; if effective harmonic interactions are employed in the CG model, the force constant of each EIC is obtained. It is important to note that the fluctuation of an EIC not only depends on its own force constant, but also on the force constants of other EIC's. This is because in typical CG models of biomolecules, the internal coordinates are interconnected. Therefore, a self-consistent solution of the effective force constants (or force fields) needs to be found, for example, by using the following recursive formula:

$$k_j^{n+1} = k_j^n + \alpha \left(\delta \mathrm{EIC}_j^n - \delta \mathrm{EIC}_j^{\mathrm{AA}} \right). \tag{4}$$

Here, k_j is the force constant of the jth EIC in a CG model if a harmonic type of interaction is used, $\delta \mathrm{EIC}_j$ is the fluctuation of the jth EIC computed from NMA, and $\delta \mathrm{EIC}_j^{\mathrm{AA}}$ is the fluctuation of the jth EIC extracted from an fined-grained simulation (in this case, an all-atom, AA, simulation). The superscript n denotes the number of iteration steps, and α is a positive numerical factor. The philosophy is that even if k_j affects the fluctuations of all other EIC's, its effect on the jth EIC should be the most significant; the interdependence of EIC's are incorporated by the iteration process. Furthermore, the allowed values of force constants are limited only to positive numbers or zero; therefore, the force constant of an EIC becomes stationary only if the computed fluctuation matches with the all-atom result or if it is decreased to zero. A force constant of zero indicates that the fluctuation of an EIC is still smaller than that in all-atom trajectories even if its force constant is reduced to zero, and such an EIC is considered redundant in a CG model and is automatically picked up during the fluctuation-matching procedure. The fluctuation-matching procedure is thus a systematic way to mimic the pattern of internal coordinate fluctuations in all-atom simulations, and the number of EIC's whose fluctuations match with those in all-atom simulations is maximized in a CG model.

3. Results

In this section, the application of the FM and fluctuation-matching coarse-graining approaches to study the biophysical properties of bioassemblies is presented. In particular, the FM method is applied to CG lipid bilayers and the fluctuation-matching method is applied to model the actin filament.

3.1 The application of force-matching to simulate dimyristoylphosphatidylcholine (DMPC) lipid bilayer

In this section, the results [9] of employing the FM method with spline interpolation and least square fit algorithm [11] to simulate a dimyristoylphosphatidylcholine (DMPC) lipid bilayer are presented. The CG force fields that are used to describe the interactions among the CG sites (water–water, water- DMPC and DMPC–DMPC) were derived using atomistic configuration and force data from a single MD simulation of the DMPC bilayer. The atomistic DMPC MD simulation consisted of 64 DMPC molecules fully hydrated by 1312 water molecules. The atomistic simulation was performed under the constant NPT ensemble. The DMPC lipid molecules were described by using a united atom force field [14]; for water molecules, the rigid TIP3P model was used [15]. The temperature of the system was kept constant at 308 K using the Nosé-Hoover thermostat with a relaxation time of 0.2 ps. The electrostatic interactions were calculated via the particle mesh Ewald summation, and a cutoff of 10 Å was used for computing the van der Waals interactions. The initial structure of the DMPC bilayer was equilibrated for 6 ns with a time

step of 2 fs. The equilibrium volume of the supercell was 107.83 nm^3 with an area per headgroup of 0.61 nm^2. In our MS-CG model, water molecules were mapped into a single interaction site associated with their geometrical center. The lipid molecules were CG as depicted in figure 1.

For the construction of configuration data for force matching, the atomistic MD simulation was carried out at the constant NVT ensemble for 400 ps with the volume set at the value obtained from the constant NPT simulation. The trajectory and force data were saved at an interval of 0.1 ps, and the total number of configurations was 4000. These trajectories of configurations and forces were then collapsed onto the CG sites, and the resulting CG trajectories were the inputs of FM. Even though some CG sites are charged (e.g. CH and PH in figure 1), it is possible to omit the Coulomb interactions in a CG model. This is because even if Coulomb interactions are involved in the CG model, the resulting effective charges on the CG sites obtained from FM are much smaller than those inferred from the atomistic charges. This is a result of screening effects from the environment (water and polar DMPC groups). Therefore, the short-ranged interactions obtained from the FM procedure effectively account for the missing Coulomb interactions. Selected pair wise CG force profiles are shown in figure 2. In a MS-CG MD simulation, the effective forces were tabulated on a fine mesh and the DL_POLY computer program [16] was used. In DL_POLY, these tabulated forces were identified as nonbonded interactions and their corresponding potentials were obtained by integrating the forces and then shifting the potential to zero at the cutoff radius. Any unevenness in the force profiles shown in figure 2 is generally caused by limited statistics but does not cause difficulty to propagate a CG MD simulation. This is due to the fact that statistical noise can be smoothed out by the spline fit.

The sites of a DMPC lipid molecule, CH, PH, GL, E1, and E2 shown in figure 1, interacted with each other through harmonic bonds. Connecting the nearest neighbor E1, E2, SM, and ST sites using harmonic bonds therefore enforced the stiffness and length of the alkane chains; the highest bond frequency in a CG lipid is about 400 cm^{-1}. The nonbonded FM (tabulated) forces are then used to describe the interactions among different lipid molecules. The CG simulation has the same system size as that in the atomistic MD simulation. The CG MD was carried out at the constant NVT ensemble with the same volume used in the atomistic MD simulation. The constant NVT conditions were adopted due to the fact that the bare CG water force field for the TIP3P model was unable to reproduce the correct water density under the constant NPT conditions even though the structural properties agree with those obtained from the atomistic MD simulation. It could be possible to perform further parameterization of the water CG force field to give a correct pure water density without changing the structural properties if simulations under the constant NPT ensemble are desired.

By using the FM procedure, the resulting CG parameters can be employed to reproduce the lipid bilayer structural properties accurately as shown in figure 3, where the selected site–site distribution functions are compared with the corresponding atomistic MD radial distributions functions (RDFs) (figure 1). The site–site RDF's that are not shown in figure 3 also agree well with the atomistic results. Despite the significant simplifications that have been made in the CG model, the observed agreement is quite good. The solvation numbers of the lipid groups in the CG simulation also agree well with

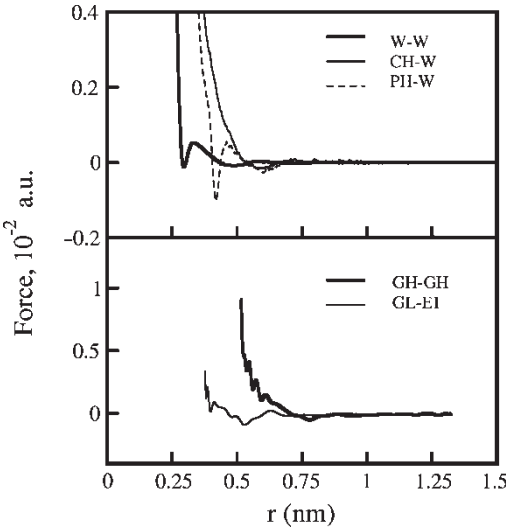

Figure 2. The effective pairwise CG forces calculated by the FM method as a function of the intersite distance between selected pairs of CG sites. The definition of CG sites are listed in figure 1 and "W" stands for water.

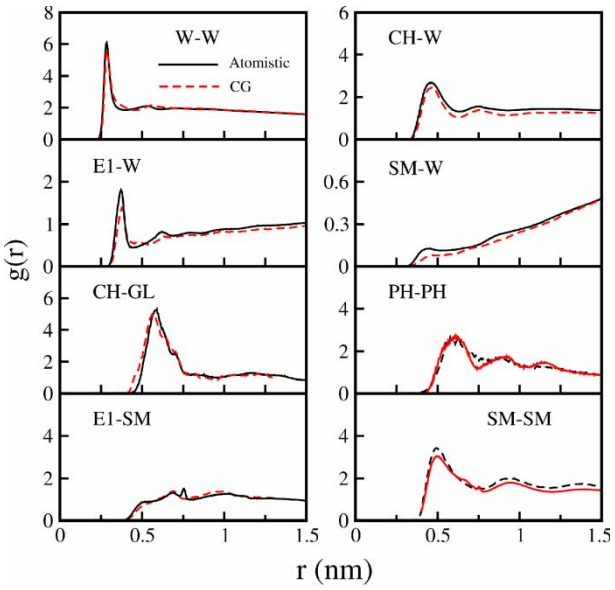

Figure 3. Comparison of selected site–site radial distribution functions obtained from atomistic and CG simulations. RDFs from atomistic MD simulations are shown in solid lines and those obtained from CG simulations are shown in dashed lines.

those of the atomistic simulation. For example, the CH site has a coordination number of 15.9 in the first solvation shell in the atomistic MD simulation and 15.6 in CG simulation. The numbers for the E1 group are 2.7 and 2.1 for the atomistic and CG simulations, respectively. The gain in speed of energy calculation in the CG simulation is 50 times compared to the atomistic simulation. However, this speed up is expected to be even greater for larger systems since the CG simulation only contains short-ranged forces; it is also expected that a significantly larger time step can be used in CG simulations.

Beyond the structural properties, it is also one of the goals of CG modeling to describe the dynamical properties in biological systems. Since CG modeling is performed on the effective free energy surface (PMF) of the CG particles, running classical MD on such CG "potentials" cannot yield any realistic dynamical information that is consistent with the First and Second Fluctuation-Dissipation Theorems. The purpose of CG classical dynamics is to explore the effective free energy surface in a more efficient fashion, but the resulting dynamics cannot reflect the actual behavior of the CG sites. Additional terms are needed in the dynamical equations used to propagate the CG sites in order to accomplish such a goal (e.g. generalized Langevin-like friction kernel and random forces). Such methodologies are currently being developed in our group.

In this section we have described a multiscale method for obtaining the effective CG force fields directly from atomistic MD simulations. The resulting CG model can reproduce the structural properties of a lipid bilayer quite accurately. The FM approach thus provides a systematic way to coarse-grain the underlying atomistic force fields with the use of spline interpolation and least square algorithms to enhance its computationally efficiency. Applications of FM to more complex biomolecular systems are currently underway.

3.2 The application of fluctuation-matching to simulate actin filaments

In this section, the results of applying the fluctuation matching method to model actin filaments are presented [10]. Actin filaments (F-actin) are the most abundant component of the cellular cytoskeleton, and they also plays critical roles in numerous processes in eukaryotic cells: cytoskeletal support, cell motility, cell division, endocytosis, and intracellular transportation [17]. Understanding how the structures and conformation of monomeric building blocks (G-actin) confer the specific properties of F-actin is thus important in the fields of molecular biology and biophysics since the functions of actin filaments are closely related to its mechanical properties such as the flexural rigidity.

F-actin is an assembly of the protein G-actin (375 residues) [18]; monomeric actin bound with ATP assembles into filaments under physiological salt concentrations. The ATPase activity of actin increases after polymerization, and the hydrolysis of bound ATP into ADP is a major force that regulates the growth dynamics of actin filaments. The dissociation of the inorganic phosphate (P_i) after ATP hydrolysis is related to the destabilization of the filament and its disassembly [18]. Atomistic MD simulations with explicit solvent indicated indicate that the conformational change of G-actin induced by ATP hydrolysis: a loop-helix transition in the DB-loop [19–21], significantly weakens the inter-monomer interactions of actin assemblies, and thus leads to a wider, shorter and more disordered filament [10]. Based on the results of atomistic simulations, CG models of F-actin can be built using the fluctuation matching procedure to further explore the elastic properties of F-actin.

Our CG model of F-actin is shown in figure 4; each CG G-actin contains four sites, and each site corresponds to one of the four subdomains of G-actin. The adenosine group of ATP or ADP is incorporated into D3 (see figure 4 for definitions) and the phosphate groups are incorporated into D4. The intramonomer interactions include three bonds, two angles, and one dihedral angle so that important slow modes of motion of G-actin, such as the propeller rotation and the open/close of the ATP cleft, are incorporated straightforwardly. For inter-monomer interactions in an F-actin, effective harmonic bonds are used. Between each pair of CG monomers, there are 16 (4×4) distinct effective harmonic bonds, and all of these bonds

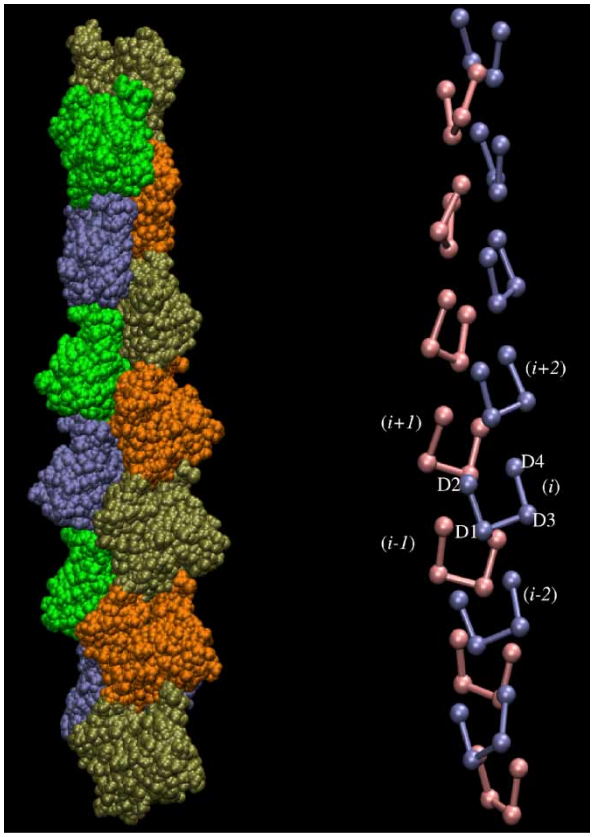

Figure 4. Atomic (left) and CG (right) representations of an actin filament. Each actin monomer has four sites and are denoted as Dn, $n = 1$–4.

are considered explicitly in the CG model. The structure of F-actin is also shown in figure 4. Based on the Holmes model from X-ray fibre diffraction, F-actin is a left-handed helix with a rotation of 166° and a rise of 27.5 Å [20]. F-actin can also be viewed as two right-handed helices wound together, and each filament is referred to as a protofilament. Each monomer in F-actin is in direct contact with four neighbors in the filament as shown in figure 4, two from the same protofilament and two from the other protofilament. It is thus obvious that effective harmonic bonds should be placed between monomers i and $i \pm 1$ in the opposite helix and i and $i \pm 2$, see figure 4. However, with reduced degrees of freedom and the simplified forms of interactions, it could be possible that additional interactions other than nearest neighbor ones are required in the CG model in order to reproduce certain properties of F-actin.

For such a CG model of F-actin, the cutoff distance, R_k, within which an effective harmonic is placed is the adjustable parameter since the force constants of EIC's are going to be determined by the fluctuation-matching procedure. Different R_k's in the CG models of F-actin are represented by the largest interval between two monomers within which effective harmonic bonds are placed, and all 16 effective bonds between any two pairs of monomers are explicitly considered. In this work, the largest interval that has been studied is 6; that is, monomer i in the filament can have effective bonds with monomers $i \pm m$, $m = 1–6$.

Since the flexural rigidity of F-actin is critical in conferring it functions, the characterization of material properties such as persistence lengths is important. The computed persistence lengths of CG models as a function of R_k are shown in figure 5. It can be seen from figure 5 that the flexural rigidity of a CG F-actin increases monotonically with R_k, the maximum interval within which a G-actin pair have effective harmonic bonds. Including effective bonds only with the nearest neighbor for a monomer with the monomers of the other protofilament, i.e. $R_k(1$, the resulting persistence lengths are 7–8 times smaller than the results of atomistic MD simulations. With minimum interactions between the two protofilaments, F-ATP has a persistence length 40% longer than that of F-ADP. This is because the loop-to-helix transition of DB-loop induced by ATP-hydrolysis weakens the interactions between the two protofilaments [22], and these changes of interactions translate into smaller values of force constants in the CG model of F-ADP.

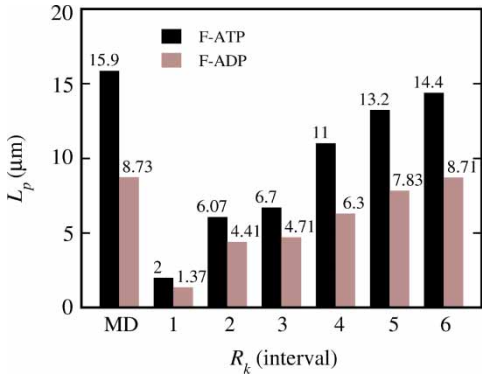

Figure 5. The persistence lengths of CG F-actin, F-ATP (black) and F-ADP (grey) as a function of R_k. The values computed from all-atom simulations are labeled as "MD".

Including effective bonds for a monomer with its nearest neighbor monomer along the same protofilament (the $i - (i \pm 2)$ interactions in figure 5), $R_k(2$, increases the persistence lengths by 3–4 times, indicating that the interactions along each protofilament are important in conferring the elasticity of an actin filament. However, the resulting flexural rigidity is still 1–2 times smaller than the atomistic results, even when all possible effective bonds among actin monomers that are in direct contact have been included. Therefore, in CG modeling where the number of degrees of freedom is reduced and the forms of interactions are simplified, criteria simply based on physical inspection may not be enough. Although at this point, the force constants could be scaled to reproduce the target values of persistence length, our goal is to directly reproduce both the fluctuations of EIC's and the flexural rigidity of F-actin.

By introducing harmonic interactions beyond the nearest monomers, the flexural rigidity of CG filaments gradually increases as shown in figure 5. With $R_k(6$, an actin monomer is connected with its first to third nearest monomers, the computed persistence length for F-ATP is 14.4 μm and for F-ADP is 8.7 μm, in good agreement with the experimental measurements (F-ATP: 13.5 μm, F-ADP: 9 μm) [23]. Figure 5 also shows that the incremental increase of L_P decays with the inclusion of longer effective bonds. This is because the force constants of effective harmonic bonds decrease with its equilibrium length. The distance dependence of the force constants will be discussed later.

In our CG models of F-actin, the force constants are uniquely determined by matching the fluctuations of EIC's in all-atom simulations. The above results suggest that instead of modifying the force constants of the effective bonds, the cutoff radius, R_k, can be used to adjust the flexural rigidity of a CG F-actin. The persistence lengths of CG F-actin with effective bonds between a monomer and its first to third nearest monomers agree with all-atom simulations and experimental measurements quantitatively. Our fluctuation-matching method is thus able to reproduce both the pattern of EIC fluctuations and the elastic properties from the fine-grained MD description.

In figure 6, the force constants of intermonomer harmonic bonds in CG F-ATP and F-ADP are plotted as a function of their equilibrium lengths. These intermonomer effective bonds are divided into two groups. The first group (G1) contains the nearest harmonic bonds, $R_k \leq 2$, and the second group (G2) contains the rest of the bonds, $R_k > 2$. In figure 6, the G1 bonds are

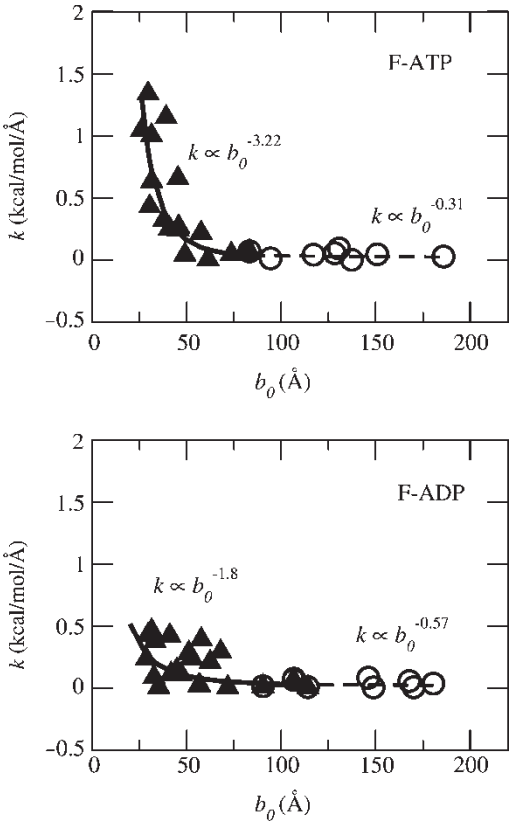

Figure 6. The force constants, k's, of the intermonomer harmonic bonds of F-ATP (left) and F-ADP (right) as a function of their equilibrium bond lengths, b_0's. The effective bond between a monomer and its nearest neighbor monomers are shown as filled upper triangles. The longer-range effective bonds are shown in empty circles.

represented as filled upper triangles, and the G2 bonds are shown as empty circles. For the G1 bonds, the force constants decay with their equilibrium length, and a power law expression was found to describe the data better than an exponentially decaying function. The orders of decay for the G1 bonds were 3.22 for F-ATP and 1.8 for F-ADP (the solid lines in figure 6). For G2 bonds, the force constants also decay with their equilibrium lengths but with smaller orders, 0.57 for F-ATP and 0.31 for F-ADP (the dash lines in figure 6).

The convergent behavior of persistence lengths as a function of R_k shown in figure 5 reflects the decay of force constants with bonds lengths as shown in figure 6. However, the decay of force constants become slow when they go beyond the nearest monomer effective bonds (75 $\sim$ 80 Å), and makes the convergence slow. An order of dependence less than one of the force constants also indicates that the effective elasticity is actually higher. This is also the reason that the addition of longer effective bonds increases the persistence lengths.

From the above result, the following strategy can be deduced for developing CG models of large bio-assemblies, such as microtubules, using harmonic type of interactions for which all-atom MD simulations may not be feasible. The specific pattern of the protein-protein interactions can be obtained by performing MD simulations on small clusters of the composing protein monomers (for example, a cluster of a tublin with its nearest neighboring monomers). Elastic properties such as the flexural rigidity is more sensitive to the effective bonds of longer lengths, and these interactions could be adjusted so that a CG model has the desired flexural rigidity.

The fluctuation-matching procedure couples the results of all-atom simulations to CG models self-consistently, and NMA can be used to provide the exact solutions of equations of motion under the harmonic approximation. Therefore, by using the CG model and NMA, one can extend the accessible time-scales by extrapolating the short-time information obtained from all-atom simulations.

In order to illustrate the above point, a computational method to compute the force-extension curves of F-actin is introduced. Harmonic bonds between the monomers at the two ends of an F-actin is used to fix the end-to-end distance by using a large value of force constant. The modified CG potential are then minimized to locate the equilibrium configuration, and NMA is then performed with the modified potential to calculate the vibrational free energy at a given end-to-end distance at 310 K. The forces are then obtained from the derivatives of the vibrational free energies at different end-to-end distances.

The force-extension curves of F-ATP and F-ADP (3 repeats, 39 monomers, and 106 nm), are shown in figure 7. The CG models used for these calculations have an R_k of 6. The maximum displacements of the end-to-end distances were chosen to be 2% of the values at equilibrium (106 nm). In Hook's regime, (positive extension and a small portion of the compression regime, $-0.4-\ -0.6$ nm), F-ATP has a stiffness of 37 pN/nm per

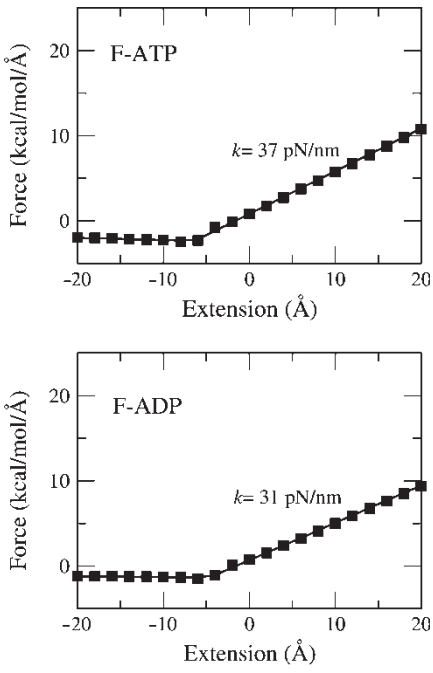

Figure 7. The force-extension curves of an F-actin with 3 repeat units (39 monomers), F-ATP (left) and F-ADP (right). The corresponding stretching stiffness per μm of a filament is labeled.

μm, and the corresponding value for F-ADP is 31 pN/nm. These values of stiffness are in reasonably good agreement with the experimentally measured value, 45 pN/nm per μm, for F-ATP labeled with PHDTMR (phalloidin-tetramethyl rhodamine), because phalloidin is known to be a structural stabilizer for the actin-filament and may increase the stiffness of F-ATP. Due to its large persistence length (elastic modulus), enthalpy dominates the contribution to the stiffness, and entropic stiffening is insignificant. After the filament is compressed beyond a displacement of 0.4–0.6 nm, little additional force is required to further bend the filament, see figure 7. For an F-actin without the presence of applied field, the slowest vibrational mode is a bending mode. Under compression, the slowest vibrational mode of a buckled filament becomes a writhing mode, whose eigenvalue is negative if the NMA is conducted without the constraining potential. Similar behavior has also been observed for an F-actin as long as 1 μm. Since it is easy to buckle an actin filament, (from figure 7, the Euler force is estimated to be 1.68 pN for a filament of a length of 1 μm, the Euler force of F-ADP is 65% that of F-ATP), the writhing mode could also be important for F-actin under a confined cellular environment in addition to the bending and twisting modes. As a result, for the generation of forces by actin filament polymerization, the writhing mode may play important roles in the interactions with the actin binding proteins for the regulation of this process.

4. Conclusions

In this work, two coarse-graining procedures, FM and fluctuation matching, that systematically incorporate the information of atomistic MD simulations into CG models are described and applied to lipid bilayers and actin filaments. Such MS-CG procedures allow the bridging of information between the fine-grained and CG levels. Structural properties, such as the pair correlation functions between different groups in a lipid bilayer, and elastic properties, such as the persistence of actin filaments, can be well captured by using such MS-CG procedures. Applications to other complex biomolecular systems are currently underway

Acknowledgements

This research was supported by grants from the National Institutes of Health (GM-063796), National Science Foundation (CHE-0218739), and a Faculty Award from the International Business Machines (IBM) Corporation. The authors also thank the National Science Foundation through TeraGrid computational resources provided by the National Center for Supercomputing Applications and the National Institutes of Health (Grant # NCRR 1 S10RR17214-01) on the Arches Metacluster, administered by the University of Utah Center for High Performance Computing.

References

[1] R. Goetz, R. Lipowsky. Computer simulations of bilayer membranes: self-assembly and interfacial tension. *J. Chem. Phys.*, **108**, 7397 (1998).
[2] S.J. Marrink, A.E. Mark. Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *J. Am. Chem. Soc.*, **125**, 15233 (2003).
[3] J.C. Shelley, M.Y. Shelley. Computer simulation of surfactant solutions. *Curr. Opin. Colloid Interface Sci.*, **5**, 101 (2000).
[4] B. Smit, K. Esselink, P.A. Hilbers, N.M. van Os, L.A.M. Rupert, I. Szleifer. Computer-simulations of surfactant self-assembly. *Langmuir*, **9**, 9 (1993).
[5] M.J. Stevens, J.H. Hoh, T.B. Woolf. Insights into the molecular mechanism of membrane fusion from simulation: evidence for the association of splayed tails. *Phys. Rev. Lett.*, **91**, 188102 (2003).
[6] J.C. Shelley, M.Y. Shelley, R.C. Reeder, S. Bandyopadhyay, M.L. Klein. Simulations of phospholipids using a coarse grain model. *J. Phys. Chem. B*, **105**, 4464 (2001).
[7] H. Meyer, O. Biermann, R. Faller, D. Reith, F. Muller-Plathe. Coarse graining of nonbonded inter-particle potentials using automatic simplex optimization to fit structural properties. *J. Chem. Phys.*, **113**, 6264 (2000).
[8] T. Murtola, E. Falck, M. Patra, M. Karttunen, I. Vattulainena. Coarse-grained model for phospholipid/cholesterol bilayer. *J. Chem. Phys.*, **121**, 9156 (2004).
[9] S. Izvekov, G.A. Voth. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B*, **109**, 2469 (2005).
[10] J.-W. Chu, G.A. Voth. Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. *Biophys. J.*, (2006) In press.
[11] S. Izvekov, M. Parrinello, C.J. Burnham, G.A. Voth. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: a new method for force-matching. *J. Chem. Phys.*, **120**, 10896 (2004).
[12] G.S. Ayton, G.A. Voth. Bridging microscopic and mesoscopic simulations of lipid bilayers. *Biophys. J.*, **83**, 3357 (2002).
[13] B.R. Brooks. Hatmonic analysis of large systems. I. Methodology. *J. Comput. Chem.*, **16**, 1522 (1995).
[14] A.M. Smondyrev, M.L. Berkowitz. United atom force field for phospholipid membranes: constant pressure molecular dynamics simulation of dipalmitoylphosphatidicholine/water system. *J. Comput. Chem.*, **20**, 531 (1999).
[15] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926 (1983).
[16] T.R. Forester, W. Smith. DL_POLY user manual. *CCLRC*, Daresbury Laboratory, Warrington, UK (1995).
[17] J. Howard. *Mechanics of Motor Proteins and the Cytoskeleton*, Sinauer Associates, Inc., Sunderland (2001).
[18] E.D. Korn. Actin polymerization and its regulation by proteins from nonmuscle cells. *Physiological Reviews*, **62**, 672 (1982).
[19] P. Graceffa, R. Dominguez. Crystal structure of monomeric actin in the ATP state – structural basis of nucleotide-dependent actin dynamics. *J. Biol. Chem.*, **278**, 34172 (2003).
[20] K.C. Holmes, D. Popp, W. Gebhard, W. Kabsch. Atomic model of the actin filament. *Nature*, **347**, 44 (1990).
[21] K.C. Holmes, D. Popp, W. Gebhard, W. Kabsch. Atomic structure of the actin: DNase I complex. *Nature*, **347**, 37 (1990).
[22] J.W. Chu, G.A. Voth. Allostery of actin filaments: molecular dynamics simulations and coarse-grained analysis. *Proc. Natl. Acad. Sci. USA.*, **102**, 13111 (2005).
[23] H. Isambert, P. Venier, A.C. Maggs, A. Fattoum, R. Kassab, D. Pantaloni, M.F. Carliers. Flexibility of actin filaments derived from thermal fluctuations. *J. Biol. Chem.*, **12**, 11437 (1995).