JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Multiscale Coarse-Graining via Normal Mode Analysis

Fei Xia and Lanyuan Lu*

School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore

Supporting Information

ABSTRACT: A multiscale coarse-graining method called the normal-mode analysis based fluctuation matching (NMA-FM) is developed for constructing coarse-grained models of biomolecular systems. In the framework of normal-mode analysis, an arbitrary fine-grained model can be systematically converted to a more coarse-grained model, while the crucial low-frequency motions of the fine-grained system are able to be reproduced in the coarse-grained modeling. The method relies on the technique of fluctuation matching that has been devised earlier for parametrizing heterogeneous elastic network models based on data from atomistic molecular dynamics simulations. The new approach is quite efficient since it avoids expensive atomistic molecular dynamics simulations and can start from already coarse-grained elastic network models. In the practical aspect, the method is suitable for conformational analyses of large biomacromolecules and calculations of mechanical properties of biomaterials, which is demonstrated by the studied systems including an amyloid dimer, lysozyme and adenylate kinase proteins, and the S2 subdomain of myosin.

1. INTRODUCTION

Atomistic molecular dynamics (MD) simulation with the conventional force fields[1−5] is frequently used in computational biophysics to investigate a wide range of phenomena involving molecules such as lipids, peptides, and proteins. The temporal and spatial scales in atomistic MD simulations are limited by computational power, which prohibits the implementation of all-atom MD simulation in the studies on many biologically relevant problems. Going beyond the limits of atomistic MD simulation, coarse-grained (CG) models can be developed to speed up the computations and explore larger scales. Numerous methods and algorithms have been devised to parametrize CG models under different circumstances. Nevertheless, the so-called multiscale CG methods have received increasing attention in recent years.[6−8]

In a multiscale CG approach, a CG model is typically parametrized based on the information produced by a model at a higher resolution, in many cases just an atomistic force field. More specifically, the CG model is often developed in a way to reproduce certain physical properties generated by the fine-grained (FG) model. Common choices of the physical properties to link the two scales include radial distribution functions[9,10] and atomistic forces.[11−13] For instance, in the multiscale coarse-graining method developed by Voth and co-workers,[13] the CG model is optimized via the fitting of underlying atomistic forces from conventional atomistic MD simulations. Alternatively, the atomistic fluctuations generated by an MD simulation can be utilized to parametrize the CG model.[14−16] This type of method involves so-called fluctuation matching, in which the calculation of the CG fluctuations is based on normal-mode analysis (NMA).

NMA has been widely applied to the studies of large magnitude conformational changes of macromolecules.[17−21] It adopts a harmonic approximation of the potential energy surface around the energy-minimized structure to describe the collective dynamics of macromolecules. Numerous successful applications of normal-mode analysis to proteins have shown that a bundle of low-frequency vibrational modes can reasonably represent the overall deformational motions of proteins at the equilibrium state.[22−26] In general, an NMA of a biological system can be carried out based on an all-atom force field. However, the computational cost of the energy minimization and the NMA process of an all-atom model are very expensive for large biological systems, and the energy-minimized structure might deviate slightly away from the experimental crystal one.

A more popular model used for protein NMA instead of all-atom models is the so-called elastic network model (ENM), including the Gaussian network model (GNM)[27,28] and the anisotropic network model (ANM).[29] ANM can be viewed as an improved version of GNM, as the latter is limited to describing only the isotropic positional fluctuations. In the ENMs for proteins, every residue is usually represented by its Cα atom. All pairs of Cα atoms connect with each other using harmonic bonds with a uniform force constant, which is inspired by the fact that a single parameter is adequate to yield the low-frequency vibrational states of folded proteins.[30] The advantage of GNM and ANM lies in that they regard the crystal structures as the energy minima for the starting points of NMA, which leads to the elimination of energy minimization and the saving of computational time. In most situations for proteins, the Cα-atom-based ENM (Cα-ENM) with harmonic potentials and uniform spring constants can reproduce the experimental B factors.[31,32] ENM can be improved further, and its development followed the idea of taking heterogeneity into account.[33−43] Very recently, Lyman et al.[16] proposed a general heterogeneous ENM, in which parameters are derived by matching the fluctuations in atomistic MD trajectories. An advantage of the method is that it can construct not only a Cα-ENM but also an arbitrarily coarse-grained ENM (CG-ENM) at any level. Because of this feature, Lyman's method is a typical multiscale

Received: July 24, 2012
Published: September 12, 2012

dx.doi.org/10.1021/ct3006387 | J. Chem. Theory Comput. 2012, 8, 4797−4806

CG method, in which an FG model is systematically converted to a CG-ENM at any resolution by MD simulation and fluctuation matching.

In this article, we propose a novel multiscale method based on NMA to construct a general CG model, which is called the normal-mode analysis-based fluctuation matching method (NMA-FM). The central idea is using the fluctuations calculated from an FG NMA to develop a CG model. Unlike the previous methods based on fluctuation matching, in our approach the fluctuations at both the FG and CG levels are computed by NMA, and therefore no FG MD simulation is needed. More specifically, given a certain system composed of many particles at a high resolution with a defined potential function, the method can calculate the fluctuations around a reference structure through NMA and construct a CG model at a lower resolution by means of fluctuation matching. Here, the FG potentials used to derive the CG potentials are not limited to the all-atom force fields. Instead, the FG potential can be any potential at any resolution such as a Cα-ENM harmonic potential or other existent CG potentials, as long as the FG potential has a higher resolution compared to the subsequently developed CG one. The commonly used CG potentials in fluctuation matching consist of harmonic springs, due to the efficient iterative algorithms developed for the parametrization of heterogeneous ENMs.[16] However, the derived CG potential can in principle have any functional form in fluctuation matching, as pointed out by Chu et al.[14,15] For simplicity, we only derive the heterogeneous CG-ENMs in all cases shown in this article.

Compared to the previous methods based on fluctuation matching and MD simulation,[14−16] NMA-FM has some unique features. In NMA-FM, the FG model is studied by NMA; thus the time-consuming FG MD simulation is avoided. Atomistic MD simulations for large biomolecular systems are usually very expensive and converge slowly, which implies that NMA-FM is advantageous in studies on large systems. In addition, inconsistency between experimental structure and simulation equilibrium structure is often found in MD simulations, which may mainly be due to force field inaccuracy. In NMA, the reference structure is mostly chosen as the experimentally available structure, and a slightly inaccurate force field will not cause a global structural deviation in the energy minimization process before NMA. As a result, NMA-FM is usually performed near an experimental structure. Another advantage of NMA-FM is that a Cα-ENM can be directly used as the FG model. By using ENM as the input, the developed CG model is supposed to be consistent with experimental information and can be linked to experimental fluctuations through B-factor data. As a trade-off for using experimental structures, NMA-FM cannot handle the situations where an experimental structure is not available. Fluctuation matching requires a well-defined reference structure of the studied system to calculate the fluctuations. Therefore, NMA-FM is particularly suitable for studies on properties of a molecular system near its equilibrium structure. Similar to NMA, NMA-FM is potentially useful for protein conformational analysis and biomechanical investigations.[44−48] The previous studies showed that the slowest motions can still be sufficiently predicted even by the low-resolution CG models,[49−51] which is also demonstrated in this article.

The rest of this paper is organized as the following: Section 2 gives the theoretical background and the numerical method. In section 3, three application examples are discussed. The first application shows the CG model of an amyloid dimer developed from the all-atom force field. The second case involves the conformational analyses of adenylate kinase and lysozyme by using the heterogeneous CG-ENMs constructed from their Cα-ENMs. The third case displays an application of NMA-FM to the calculation of biomechanical properties of the S2 subdomain of myosin. Finally, the conclusions and the future directions are provided in section 4.

2. THEORY AND COMPUTATIONAL DETAILS

2.1. NMA-FM Method. The NMA-FM method aims to develop a CG model based on an existent FG potential, given a known reference structure of the molecular system that is usually from experimental data. The approach follows the idea of fluctuation matching, in which the fluctuations around the reference structure are the key properties to connect the CG and FG models. These fluctuations are calculated separately for some effective internal coordinates (EICs), based on which the CG potential is built, namely,[15]

$$\Delta \mathrm{EIC}_k^2 = (\mathrm{EIC}_k - \overline{\mathrm{EIC}_k})^2 \tag{1}$$

Here, $\Delta \mathrm{EIC}_k{}^2$ is the fluctuation for the kth EIC, and $\overline{\mathrm{EIC}_k}$ is the equilibrium value that is usually calculated on the basis of the experimental structure. For instance, in a CG-ENM each harmonic spring corresponds to an EIC r_k, which is the pairwise distance for the kth spring. Similarly, EICs can be constructed for arbitrary CG potential functional forms, such as conventionally used angle, dihedral, and nonbonded interactions, and NMA-FM in principle can be applied for developing CG force fields containing various potential functions.

The inputs of ENM include an FG potential U_{FG} and an experimental structure of the studied system. Typically, the resolution of the experimental structure allows an NMA using the structure and U_{FG}. If the experimental structure has a resolution lower than that of U_{FG}, a standard molecular docking needs to be performed to obtain the FG structure for NMA. By the NMA at the FG level, a number of eigenvectors and eigenvalues corresponding to the normal modes of the system are calculated. Sometimes, only some important modes related to the slow motions are kept for computing the fluctuations. The next step is to construct the functional forms of the CG potential U_{CG} and the EICs, which involves a mapping procedure that is commonly used in multiscale CG methods.[13] In this work, we directly divide the target systems into the CG sites according to their different domain functions and use the center of mass locations as the coordinates of the sites. The normal modes calculated at the FG level are then coarse-grained and projected on the EICs to compute the fluctuation for each EIC. The details regarding the NMA and the projection are given later in this article. An initial guess of U_{CG} is subsequently used to conduct another NMA at the CG level, in which the CG normal models are projected on the EICs to obtain the CG fluctuations. An update of U_{CG} needs to be determined on the basis of the difference between the FG and CG fluctuations that are projected on the same set of EICs. The update procedure is typically iterative, which ends when the difference between the FG and CG fluctuations is minimized. Mathematically, the final U_{CG} in NMA-FM minimizes the residual in eq 2:

$$\chi^2 = \sum_k (\Delta \mathrm{EIC}_{k,\mathrm{FG}}^2 - \Delta \mathrm{EIC}_{k,\mathrm{CG}}^2)^2 \tag{2}$$

in which the two terms in the summation represent the fluctuations on the kth EIC generated by the FG and CG normal modes, respectively. Alternatively, threshold values can also be assigned to individual EICs to indicate the convergence of the CG potential parameters.

On the basis of the above discussion, the following algorithm can be designed for NMA-FM:

(1) Perform an NMA using U_{FG} based on the input structure.
(2) Determine the functional form of U_{CG} and the EICs.
(3) Project the normal modes generated by U_{FG} on the EICs to obtain the FG fluctuations.
(4) Determine an initial guess of the parameters of U_{CG}.
(5) Perform an NMA using U_{CG} and the input structure.
(6) Project the normal modes generated by U_{CG} on the EICs to obtain the CG fluctuations.
(7) Compare the FG and CG fluctuations, and update the parameters in U_{CG} accordingly.
(8) Repeat steps 5−7 until the residual defined in eq 2 is below a predefined threshold.

The final U_{CG} optimized in NMA-FM reproduces the FG fluctuations that are projected on the EICs, which indicates that the FG and CG potentials generate similar patterns of fluctuations at the CG level. Since the fluctuations are projected by the normal modes, the CG potential developed by NMA-FM is expected to reflect the critical vibrational motions of the system, as we show in the application cases.

While the input FG potential U_{FG} can be an arbitrary atomistic or CG potential as long as it is more fine-grained than U_{CG} developed later, two types of FG potentials are of particular interest based on the practical considerations. The first type of FG potentials consists of the conventional atomistic force fields. Nowadays, atomistic force fields are available for almost all types of biomolecules. However, atomistic MD simulations and NMA are known to be very expensive for many large biomolecular systems, which invokes the motivation for coarse-graining. In a standard multiscale modeling strategy, the CG modeling for a small system is performed, and the results can be used for much larger systems that contain the repeated units as the small system. Thus, NMA-FM is able to systematically convert conventional atomistic models to CG models for various biologically relevant systems. The second type of input potentials includes the Cα-ENMs that are widely used to study biomacromolecules. ENMs are parametrized on the basis of experimental structures and B-factors, and therefore it avoids the inaccuracy often found in atomistic force fields. In NMA-FM, it is convenient to use a Cα-level ENM as the input FG potential and develop a more coarse-grained potential to reach larger scales.

In principle, there is also no limitation on the functional form of the output CG potential. However, finding an efficient algorithm for the updates of U_{CG} in step 7 is not trivial in most cases. A convenient choice of U_{CG} is a heterogeneous ENM, and the update scheme developed by Lyman et al.[16] can be implemented. In this article, we restrict the output CG potentials to be heterogeneous ENMs. As a result, the EICs correspond to the pairwise distances r_{ij} defined by the harmonic oscillator between two CG sites i and j. A universal CG potential updating algorithm for other commonly used bonded and nonbonded functional forms is challenging. For very simple CG models, methods based on trial-and-error or parameter scan may be implemented. Note that the iterative procedure in NMA-FM involves only relatively fast NMA calculations at the CG level. Therefore, many numerical methods for global optimization can be in principle exploited.

2.2. Details of NMA. All energy minimizations and normal-mode analyses of the systems are carried out with a double precision version of Gromacs 4.5.5.[52,53] After energy minimization, the Hessian matrix of a system with N particles can be constructed and diagonalized to generate the $3N$ eigenvalues and eigenvectors, among which the first six modes correspond to the translational and rotational motions while the other $3N - 6$ modes correspond to the vibrational motions. The mean square fluctuations of particles are readily calculated once the eigenvectors and eigenvalues of the Hessian are known, according to the analytical expression of eq 3:[54]

$$\langle \Delta \mathbf{R}_i^2 \rangle = \frac{k_B T}{m_i} \sum_h^{\text{modes}} \frac{\mathbf{U}_{ih}^2}{\omega_h^2} \tag{3}$$

where $\Delta \mathbf{R}_i$ and m_i are the displacement and the mass of the ith particle, k_B is the Boltzmann's constant, T is the temperature, and $\mathbf{U}_{ih}$ is the ith three-dimensional subvector of the hth eigenvector with the frequency ω_h, which is the projection of the eigenvector on particle i. In principle, the summation in eq 3 includes all modes. Nevertheless, in NMA's practice, it is common to select a number of low frequency modes in the calculation of the displacement fluctuations, to better reproduce the experimental B factors.[21,23,55,56] A parametrization of the FG model is performed during the B-factor fitting when a Cα-ENM is chosen as the FG potential.

Once the FG modes are obtained via NMA, they need to be projected on the EICs (r_{ij} for the harmonic bonds in a CG-ENM) to calculate the fluctuations. This is achieved by first coarse-graining the FG modes and subsequently projecting them on the EICs. To coarse-grain the modes, each CG site is assigned at the center of mass location, and each eigenvector of size $3N$ is converted to a "CG" eigenvector of size $3N_{CG}$, where N_{CG} is the total number of CG sites. In the hth FG eigenvector, each FG particle is related to a three-dimensional subvector $\mathbf{U}_{ih}$ that corresponds to the three Cartesian coordinates of particle i. In the conversion of the FG eigenvector, the mass-weighted average of the three-dimensional subvectors from the FG particles forming a CG site is calculated, and the resultant three-dimensional subvector becomes the part of the CG eigenvector corresponding to the CG site, namely,

$$\mathbf{U}_{Ih} = \frac{1}{\sum_i^{i\in I} m_i} \sum_i^{i\in I} m_i \mathbf{U}_{ih} \tag{4}$$

Here, $\mathbf{U}_{Ih}$ is the CG projection of the hth eigenvector on the CG site I. The final CG eigenvector is composed of the CG subvectors, and the conversion is done for all of the FG eigenvectors.

The projections of fluctuations $\Delta r_{ij,FG}^2$ along all pairs of bond vectors between the CG sites are calculated using the converted CG eigenvectors,[16] where the subscript ij means that the bond fluctuation is between the ith and the jth CG sites, and the subscript FG indicates that the fluctuations at the CG resolution are calculated based on the FG potential. These fluctuations are then designated as the target for the subsequent fluctuation matching. In some cases, it is possible to include only a number of low-frequency modes from the FG model in the projection, which means that we expect the developed CG

model to reproduce the properties based on the selected important modes.

The NMA of the CG model is very similar to the FG NMA. As a result, the projections of fluctuations $\Delta r_{ij,\mathrm{CG}}{}^2$ are computed on the basis of the CG potential. The CG NMA needs to be conducted multiple times during the iterative refinement of the CG model parameters.

2.3. Potential Updating Algorithm. As already mentioned, the CG potentials in this work are heterogeneous CG-ENMs. The spring constants of the heterogeneous CG-ENM are iteratively refined by using the algorithm developed by Lyman et al.[16] For the readers' convenience, we outline the numerical procedure briefly as below:

(1) Connect all pairs of CG sites with an initial uniform spring constant k_{ij} to construct the CG-ENM.
(2) Perform an NMA, and project the fluctuations on the bonds between every pair of CG sites to compute $\Delta r_{ij,\mathrm{CG}}{}^2$.
(3) Update all spring constants simultaneously according to eq 5:

$$\frac{1}{k_{ij}^{n+1}} = \frac{1}{k_{ij}^{n}} - \epsilon(\Delta r_{ij,\mathrm{FG}}^2 - \Delta r_{ij,\mathrm{CG}}^2) \tag{5}$$

where n labels the iteration number, and the parameter ϵ controls the size of the steps in the optimization of the spring constants. The algorithm proceeds to repeat the steps 1−3, until the difference between the bond fluctuations is less than a predefined threshold value σ. In practice, the threshold is defined as the relative deviation of the bond fluctuations, calculated as $\sigma = (|\Delta r_{ij,\mathrm{FG}}| - |\Delta r_{ij,\mathrm{CG}}|)/(|\Delta r_{ij,\mathrm{FG}}|)$. If all of the calculated σ values are lower than 10^{-3}, the convergence for the spring constants is reached and the iteration process stops. In practice, the spring constants in all four of the CG-ENM systems in this article quickly reached the convergence within 1000 iterations by using the algorithm above implemented in a C script.

2.4. All-Atom Model. The initial atom coordinates of the amyloid dimer were taken from the crystal structure deposited in the protein data bank (PDB ID: 3OW9).[57] The all-atom model of the amyloid dimer was then constructed with the OPLS force field,[4] and the structure was optimized to search the energy minimum using the conjugated gradient algorithm.[58] All electrostatic and van der Waals interactions between atomistic pairs were calculated within a cutoff value 1.0 nm. The solvent effect was considered implicitly using the generalized Born formalism, where the Born radii were calculated using the Still method[59] and the dielectric constant was set to 80 to represent the water environment. The conjugated gradient algorithm stopped searching when the change of the total energy of the system was less than 10^{-6} kJ/mol and the maximal force on every atom was less than 2 pN.

2.5. Elastic Network Model. The total potential E of a homogeneous ENM is expressed as the sum of harmonic interactions as shown in eq 6:

$$E = \sum_{i}^{N}\sum_{j\neq i}^{N} \frac{1}{2}k(r_{ij} - \overline{r}_{ij})^2 \cdot H(r_c - \overline{r}_{ij}) \tag{6}$$

where k is the harmonic force constant, N is the total number of particles, r_{ij} is the distance between particles i and j, $\overline{r}_{ij}$ is the equilibrium distance between particles i and j in the crystal structure, r_c is the cutoff distance, and $H(x)$ is the Heaviside unit step function. $H(x) = 0$ if $x < 0$; otherwise, $H(x) = 1$. The cutoff r_c needs to be carefully chosen in order to reproduce the experimental B factors with high accuracy. One can adjust the parameters of the ENM to obtain the best fittings for the target systems. However, just for simplicity in this work, we used a uniform force constant 0.5 kcal/mol/Å^2 and a cutoff value 9 Å for the Cα-ENMs of adenylate kinase, lysozyme, and the S2 subdomain systems, since our preliminary tests found that this suit of settings could reproduce the experimental B factors reasonably well for all three systems. We believe that more accurate parameters derived from other sophisticated methods[37−40] will not affect our main conclusions in this study. Heterogeneous ENMs are implemented as the CG models, in which the spring constant varies from pair to pair.

For the nonequilibrium simulations[60−63] of the S2 subdomain of myosin, the constant forces ranging from 100 to 300 pN were applied to the two terminals of the CG-ENM, and the protein domain was pulled along the two opposite directions along the two terminals. For each force value, a 10 ns trajectory of Brownian dynamics was generated using Gromacs for the analyses.

3. RESULTS AND DISCUSSION

3.1. Amyloid Dimer. In this case, we show how to develop a CG-ENM from an all-atom force field using the NMA-FM method. The mechanical properties of amyloid fibrils and their biological functions related with Alzheimer's disease have been extensively studied in recent years.[64−67] The crystal structure of the amyloid dimer has two chains in the same sequence KLVFFA, with a total of 12 residues. After the energy minimization using its all-atom model, we performed the NMA and calculated the theoretical B factors for all heavy atoms. The theoretical B factors were evaluated on the basis of the modes with frequencies less than 30 cm^{-1}.[21,23,55,56] Figure 1 shows the comparison of the experimental and theoretical B factors for the heavy atoms in the amyloid dimer. The overall fitting between the red and black curves is reasonably good, especially for the heavy atoms on the backbone with the indices 1−46 and except that several heavy atoms on the side chains have larger deviations. The large discrepancy on the side chain atoms may be due to the force field inaccuracy and the

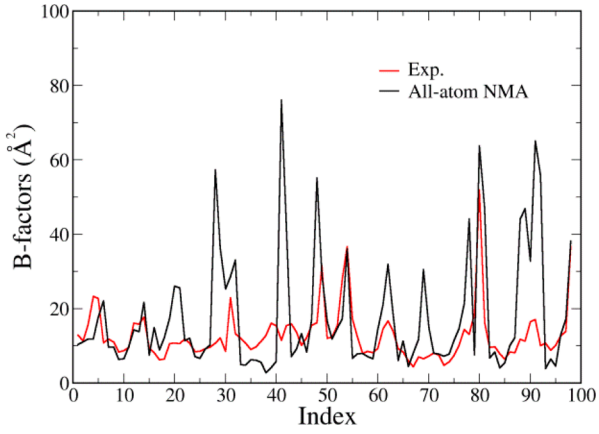

Figure 1. The comparison of experimental and theoretical B factors of the heavy atoms in the amyloid dimer. The red line denotes the crystal B factors of all heavy atoms (PDB ID: 3OW9) while the black line denotes the theoretical ones calculated from NMA. The indices 1−98 correspond to the atom IDs of heavy atoms in the PDB file 3OW9.

environmental differences between the experiment and the computation. The comparison between the experimental and theoretical B factors based on the all-atom model is similar to the previous results for alpha lytic protease.[21]

To construct the CG-ENM for the amyloid dimer, the 12 residues were mapped into four CG sites according to their sequence, namely, CG1(1−3), CG2(4−6), CG3(7−9), and CG4(10−12), as shown in Figure 2. The fluctuations along the

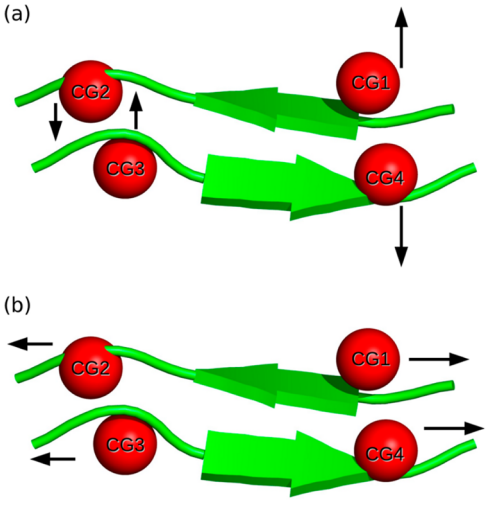

Figure 2. The cartoon representations of the all-atom model of the amyloid dimer. The amyloid dimer is mapped into four CG sites, CG1(1−3), CG2(4−6), CG3(7−9), and CG4(10−12) according to its residue sequence. (a) The arrows show the staggered bending motion of the CG sites calculated by the NMA of the CG-ENM. (b) The arrows show the stretching motion of the CG sites calculated by the NMA of the CG-ENM.

directions of bond vectors between all CG sites were then calculated by the NMA of the all-atom model. The spring constants between the CG sites were obtained by using the NMA-FM method and listed in Table S1 (Supporting Information). For the CG-ENM of the amyloid dimer, the spring constants converged quickly within a few iterations. With the derived spring constants, we performed the NMA and compared the low-frequency motions of the CG-ENM with those of the all-atom model. As for the normal modes of the amyloid dimer, we are mostly concerned with the bending and stretching motions closely associated with their biomechanical properties.[66,67] The lowest-frequency mode based on the all-atom model indicates that the terminals of the two beta sheets in the dimer follow a staggered bending motion of up and down (Figure S2, Supporting Information), with the calculated frequency 1.2 cm^{-1}. Meanwhile, it is found that the lowest mode of the CG-ENM represents a similar bending motion (Figure S3, Supporting Information) to that of the all-atom model, as shown in Figure 2a. In addition, the results of the all-atom model reveal that vibration mode 9 exhibits the slight stretching motion along its backbone, while mode 8 of the CG-ENM corresponds to the same stretching motion, as shown in Figure 2b. Our comparison of the NMA results for the all-atom model and the CG-ENM of the amyloid dimer indicates that the CG-ENM obtained using the NMA-FM method can correctly reflect the crucial low-frequency modes generated by the all-atom model. Thus, it is possible to study larger biological systems with NMA-FM, using atomistic force fields as inputs. In the following subsection, we present the results on two protein systems in which the conformational motions are investigated with the heterogeneous CG-ENMs using the Cα-ENMs as the inputs.

3.2. Lysozyme and Adenylate Kinase. Lysozyme is an important protein whose conformations have been widely studied.[68−71] The previous domain analysis based on the X-ray structure and the essential dynamics of lysozyme revealed that it has two dominant hinge-bending motions.[72] According to the motions along the eigenvectors of the hinge-bending mode, lysozyme can be divided into the functional domain A (residues 1−66) and the domain B (residues 81−164). The residues 67−80, corresponding to the interdomain helix in lysozyme, form the dynamic link between domain A and B for the hinge-bending motions. To explore the conformational motions of lysozyme, a Cα-ENM was constructed based on the Cα atom coordinates extracted from the high resolution X-ray structure of the wide-type lysozyme (PDB ID: 2LZM).[73] In order to validate the parameters of the spring constants and the cutoff value used for Cα-ENM, we compared the experimental and theoretical B factors of lysozyme. In Figure 3a, it can be seen that the theoretical B factors calculated from the NMA of the Cα-ENM reproduce the crystal fluctuations of the Cα atoms reasonably well.

In order to study the conformational motions with a CG-ENM, we coarse-grained the Cα-ENM of lysozyme directly according to its functional domains. As shown in Figure 4,

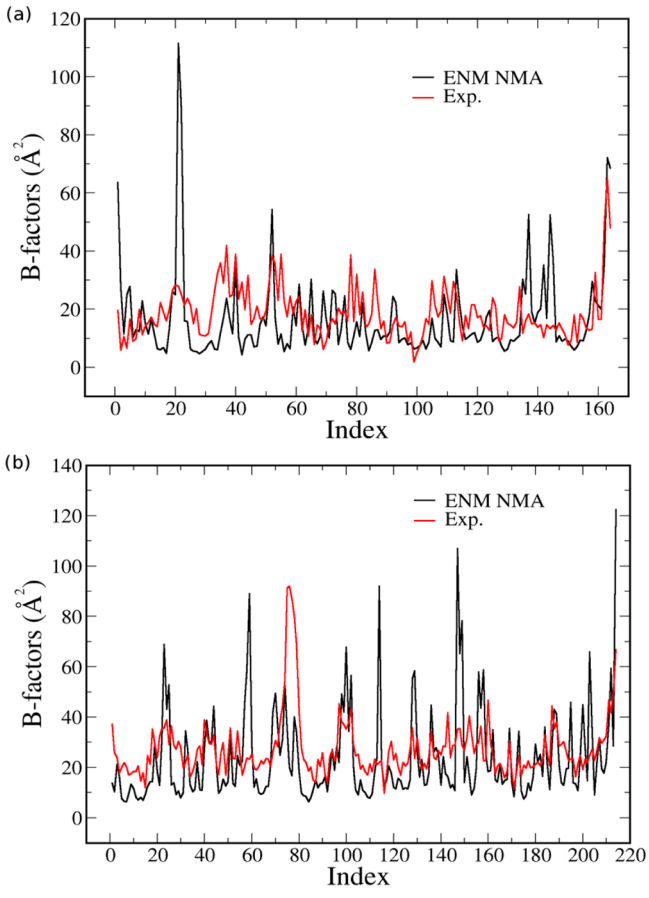

Figure 3. The comparison of the experimental (Exp.) crystal B factors colored in red and theoretical B factors in black calculated from the NMA of the Cα-ENM. (a) The comparison for lysozyme. (b) The comparison for adenylate kinase.

dx.doi.org/10.1021/ct3006387 | J. Chem. Theory Comput. 2012, 8, 4797−4806

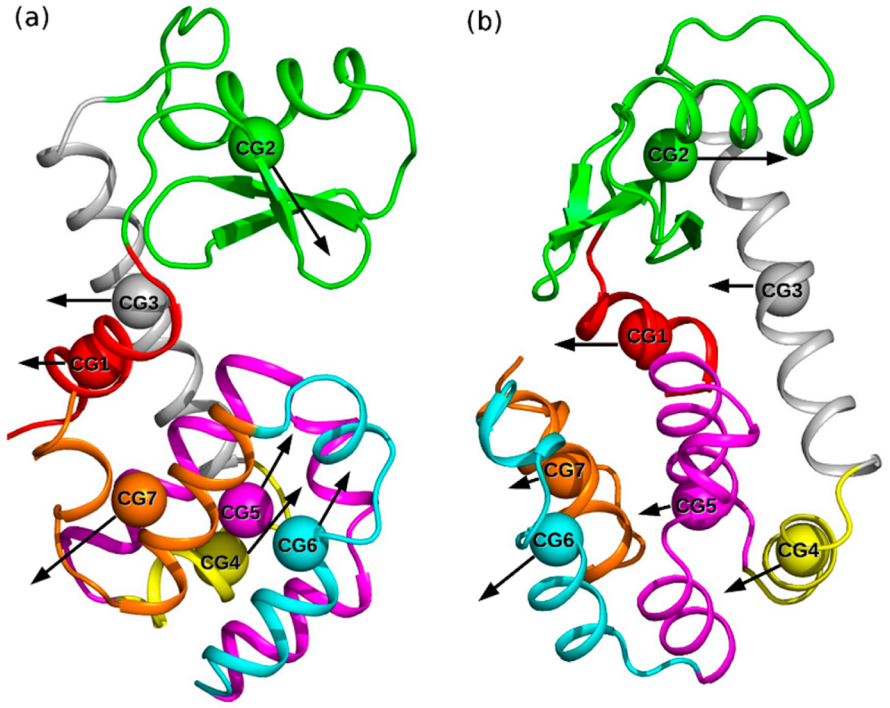

Figure 4. The cartoon representations of the Cα-ENM and the CG division of lysozyme. Lysozyme is mapped into seven beads according to the sequence of the Cα atoms, namely, CG1(1−12), CG2(13−58), CG3(59−80), CG4(81−91), CG5(92−123), CG6(124−143), and CG7(144−164). The coarse-grained beads CG1∼CG7 are colored in red, green, gray, yellow, magenta, cyan, and orange, respectively. (a) The side view of the lysozyme shows a closure bending motion between domain A(CG1∼CG2) and domain B(CG4∼CG7), where the arrows indicate the direction of the closure motion. (b) The front view of the lysozyme shows a twisting motion between the two domains A and B, where the arrows indicate the deformational direction.

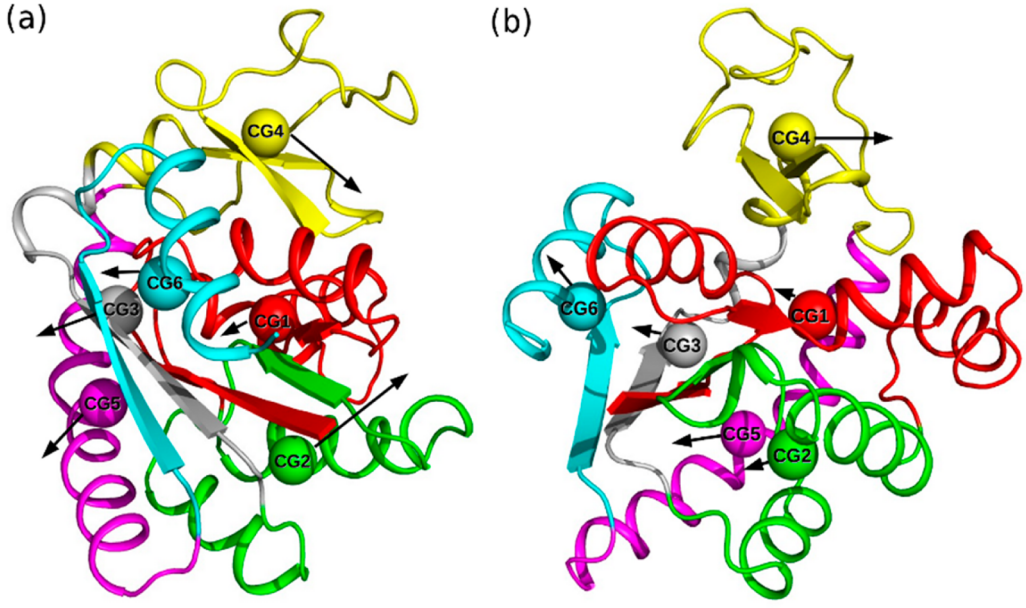

Figure 5. The cartoon representations of the Cα-ENM and the CG division of adenylate kinase. The whole system is mapped into six beads according to the sequence of Cα atoms, namely, CG1(1−59), CG2(60−102), CG3(103−117), CG4(118−160), CG5(161−190), and CG6(191−214). The beads CG1∼CG6 are colored in red, green, gray, yellow, magenta, and cyan, respectively. (a) The side view of the adenylate kinase shows a closure bending motion between CG4 and CG2, where the arrows indicate the closure direction. (b) The front view of the adenylate kinase shows a twisting motion between CG4 and other beads, where the arrows indicate the deformational direction.

lysozyme is coarse-grained into seven beads CG1∼CG7. Domain A of lysozyme consists of beads CG1∼CG2, while domain B is composed of beads CG4∼CG7. The bead CG3 serves as the dynamic link between the two domains A and B. The spring constants for the CG-ENM containing seven sites were parametrized using the NMA-FM method, as shown in Table S2 (Supporting Information). Subsequently, we performed an NMA and compared the low-frequency modes

of lysozyme from the homogeneous Cα-ENM and the heterogeneous CG-ENM. The NMA of the Cα-ENM indicates that the lowest-frequency vibrational mode corresponds to a closure motion between domain A and domain B (Figure S4, Supporting Information), especially for their helix parts. The calculated frequency for the lowest closure mode is 2.5 cm^{-1}, in good agreement with the previous result of 3.6 cm^{-1} by Brooks and Karplus.[70] The second lowest-frequency mode of lysozyme is a twisting motion between domains A and B, with a calculated frequency of 3.1 cm^{-1}. The two lowest-frequency modes are closely associated with the catalytic function of lysozyme as interacting with the substrates. The NMA of the CG-ENM reveals that the two lowest-frequency hinge-bending motions from the Cα-ENM are correctly reproduced. As shown in Figure 4a, the lowest frequency mode of the CG-ENM represents the closure motion of lysozyme, where a stretching motion between bead CG2 and beads CG4∼CG5 can be clearly distinguished (Figure S5, Supporting Information). The calculated frequency of the stretching mode is 2.7 cm^{-1}, in good agreement with the corresponding 2.5 cm^{-1} from the Cα-ENM of lysozyme. The second lowest-frequency mode in Figure 4b corresponds to the twisting motion between beads CG1∼CG2 and CG4∼CG7, with a value of 4.7 cm^{-1} similar to the 3.1 cm^{-1} from the Cα-ENM. Thus, it is evident that the current CG-ENM can not only reproduce the crucial low-frequency modes from the Cα-ENM but also yield the proper eigenfrequencies that reflect the strength of the fluctuations.

The second case involves the conformational analysis of adenylate kinase. Adenylate kinase is a phosphotransferase enzyme that catalyzes the transfer of ATP to AMP in the cell.[74−77] The structure of adenylate kinase is composed of three domains, the core domain with residues 1−29, 68−117, and 161−214; the ATP binding domain of residues 118−167; and the monophosphate binding domain of residues 30−67.[75] Adenylate kinase is capable of transiting between the conformations of open and closed. In this case, the adenylate kinase in a closed state serves as the target for conformational analysis. The Cα-ENM system was constructed from the crystal structure (PDB ID: 1AKE).[78] The comparison of experimental and theoretical B factors in Figure 3b manifests that the adopted parameters for the Cα-ENM of adenylate kinase are able to reproduce reasonably the experimental B factors. Similar to the CG division of lysozyme, we map adenylate kinase into six beads CG1∼CG6, as shown in Figure 5. The spring constants of the CG-ENM were derived by using the NMA-FM method and are listed in Table S3 (Supporting Information). The NMA results of the Cα-ENM indicate that the adenylate kinase possesses two hinge-bending motions analogous to lysozyme. The lowest-frequency mode of the Cα-ENM corresponds to a closure bending motion between the ATP binding domain (118−167) and the loop region (74−80) (Figure S6, Supporting Information), with a frequency of 3.8 cm^{-1}. Figure 4a shows the lowest-frequency mode obtained from the CG-ENM of the adenylate kinase, where the ATP binding domain is almost represented by the bead CG4(118−160), and the loop region is incorporated into bead CG2 (60−102). The lowest vibrational mode clearly exhibits a closure motion between the CG4 and CG2 (Figure S7, Supporting Information), with a calculated frequency of 3.2 cm^{-1}. Figure 4b also shows the second lowest-frequency mode of the CG-ENM, indicating a twisting motion between the CG4 and other beads. The calculated frequency of the twisting mode is 5.5 cm^{-1}, close to the 4.5 cm^{-1} of the corresponding mode of the Cα-ENM. The conformational analysis of the adenylate kinase further demonstrates that the CG-ENM model can reproduce the crucial low-frequency modes that reflect the primary fluctuations of biological systems, even with the reduced degrees of freedom. The frequencies of the hinge-bending motions calculated from the CG-ENM are nearly identical to these from the Cα-ENM, contributing to the good reproducibility of the crucial low-frequency modes.

3.3. S2 Subdomain of Myosin. The heterogeneous CG-ENMs developed by the NMA-FM method can be also applied to the studies on mechanical properties of biological systems. In this part, we show how to calculate the stretching stiffness of the S2 subdomain of myosin using a CG-ENM. The reason for choosing the S2 subdomain myosin is its biological importance for muscle contraction.[79−82] So far, only the 10-nm-long coiled coils of the S2 subdomain of myosin are available in the crystal structures. We first constructed the Cα-ENM of the S2 subdomain from the crystal structure of S2 subdomain (PDB ID: 1NKN).[83] Figure 6 shows the comparison of experimental and theoretical B factors for the Cα atoms of the S2 subdomain. Using the same parameters of Cα-ENMs as those for lysozyme and adenylate kinase, we obtain a reasonable distribution of B factors comparable to the experimental ones. To derive the CG-ENM, the whole S2 subdomain with the 150 Cα atoms is mapped into six beads, namely, CG1(1), CG2(2−60), CG3(61−74), CG4(75−90), CG5(149), and CG6(150), as shown in Figure 7. The beads CG1 and CG6 are set as the terminal beads responsible for the measurement of the end-to-end distance of the coiled coil. The derived spring constants of the CG-ENM are displayed in Table S4 (Supporting Information). The previous NMA investigation of the S2 subdomain based on an all-atom model[84] found that the lowest-frequency mode is a bending motion with a frequency of 0.7 cm^{-1} while mode 24 corresponds to a stretching motion with a frequency of 6.9 cm^{-1}. Our NMA results based on the Cα-ENM also indicate that the lowest-frequency mode is a bending motion (Figure S8, Supporting Information), and mode 18 corresponds to a stretching motion, whose frequencies are 0.4 and 4.9 cm^{-1}, respectively. Hence, our results are consistent with those of the previous all-atom calculations. Moreover, the NMA of the CG-ENM indicates that the crucial bending (Figure S9, Supporting Information) and stretching motions

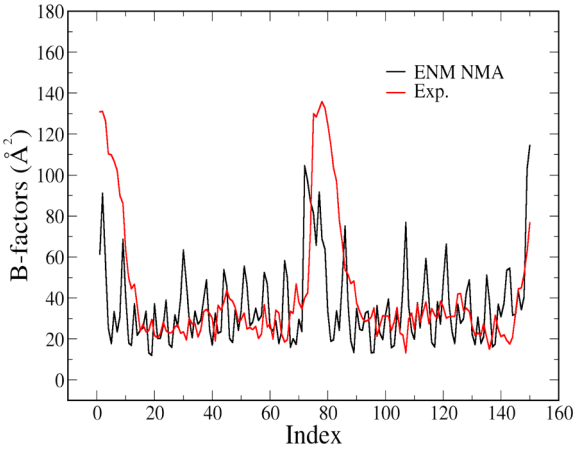

Figure 6. The comparison of the experimental (Exp.) B factors of the S2 subdomain of myosin and the theoretical ones calculated from the NMA of the Cα-ENM.

dx.doi.org/10.1021/ct3006387 | J. Chem. Theory Comput. 2012, 8, 4797−4806

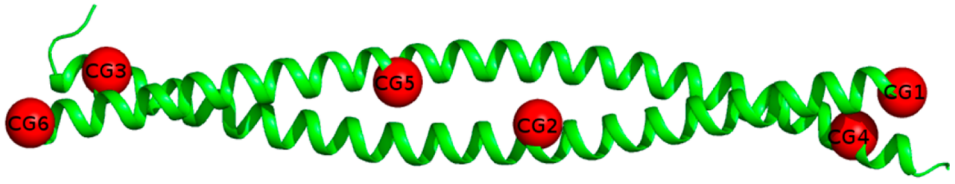

Figure 7. Cartoon representation of the Cα-ENM of the S2 subdomain of myosin. The Cα atoms are coarse-grained as six beads, CG1(1), CG2(2−60), CG3(61−74), CG4(75−90), CG5(149), and CG6(150), according to their residue sequence.

are well produced in modes 7 and 12, respectively, with frequencies of 0.2 and 3.0 cm^{-1}. Finally, we carried out the nonequilibrium Brownian dynamics simulations using the CG-ENM of the S2 subdomain. The constant forces ranging from 100−300 pN were applied to the two terminal beads CG1 and CG6, pulling them in the opposite directions along their connecting vector. The relationship between the forces and extensions is plotted in Figure 8. A linear fitting of the red dots

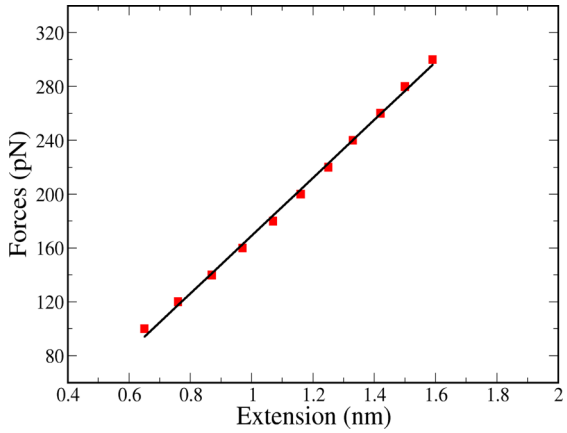

Figure 8. Relationship between the forces and the extensions, presented as the red dots. The forces range from 100−300 pN at an interval of 20 pN. The Brownian dynamics simulations were carried out at different forces to generate the 10 ns trajectories for analyses. The extension is calculated as the difference between the average end-to-end distance of the force-clamped trajectories and the original length 10.5 nm between beads CG1 and CG6 shown in Figure 6.

in Figure 8 gives rise to a slope of 215 pN/nm, which accounts for the stretching stiffness of the S2 subdomain fragment. Scaling the result to a long S2 of 65 nm,[84] the stretching stiffness is approximately 35 pN/nm, falling in the experimental range of 32−80 pN.[85,86] More accurate result may require a more precise parametrization of the Cα-ENM. Heterogeneous ENM is particularly useful for biomechanical calculations because mechanical properties of biological systems such as bending stiffness, stretching stiffness, and twisting stiffness are intrinsically heterogeneous in nature. The current method is suitable for the study of mechanical properties of very large systems, especially the systems consisting of repeated units such as the vimentin proteins.[87,88]

CONCLUSION

We present a new multiscale CG method, in which a CG model is parametrized on the basis of a FG model and NMA. Currently, the output CG model is confined to be a heterogeneous ENM in order to implement the previously devised algorithm. We show how to derive the heterogeneous CG-ENMs from either an all-atom model of an amyloid dimer or Cα-ENMs of proteins. The first advantage of the NMA-FM method is its high computational efficiency, even for large systems. It only needs an NMA based on the FG model and avoids the time-consuming MD simulations. The calculations are further speeded up by using Cα-ENMs as the input FG models. The second feature of the NMA-FM method is that the derived CG-ENM can reproduce the crucial low-frequency modes similar to the counterparts at the high resolution, which is supported by the conformational analyses of lysozyme and adenylate kinase. For both lysozyme and adenylate kinase, their hinge-bending motions are correctly characterized by their corresponding CG-ENMs. Third, the NMA-FM method can be applied to the evaluation of mechanical properties of biological systems. We show how to calculate the stretching stiffness of the S2 subdomain of myosin by constructing the heterogeneous CG-ENM. Practically, the method can be potentially applied to very large biomolecular systems, especially those composed of repeated units. Since the method is able to produce CG models at arbitrary resolutions, very aggressively coarse-grained models can be built by the method, which may bridge the gap between the particle-based modeling and the macroscopic mechanical computations traditionally based on continuum mechanics and finite elements.[89,90] Applications of the NMA-FM method to the biomechanics calculations ought to be conducted in the future.

The current NMA-FM method can be improved further in several ways. One direction is to take the fluctuation correlations between the CG interactions into account in the iterative matching algorithm. At present in the CG potential updating cycle, a spring force constant is only related to the fluctuation of the pairwise distance between the two sites connected by the spring, whereas the correlations between different harmonic bonds are not considered in the algorithm. The deficiency of the algorithm might lead to convergence problems for some complicated systems. In addition, conventional force field terms can be introduced into the CG model, which requires new updating algorithms for the potentials other than harmonic bonds. Hence, our method may be improved for biomechanical calculations by considering the nonharmonic behaviors of interactions, which is believed to be important for the heterogeneous biomaterials far away from the equilibrium state.

ASSOCIATED CONTENT

Supporting Information

Tables of converged spring constants, a plot of deviations of theoretical B-factors from the experimental ones, figures illustrating the lowest-frequency mode for each of the four FG/CG systems, and movies in GIF format corresponding to the eight figures. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: LYLU@ntu.edu.sg.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Prof. Zhiyong Zhang for some valuable discussions regarding NMA. This research was supported by a start-up grant provided by Nanyang Technological University, and part of the computational resource was from an IDA cloud computing grant.

REFERENCES

(1) Brooks, B. R.; Bruccoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(2) MacKerel, A. D., Jr.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; John Wiley & Sons: Chichester, U. K., 1998; Vol. *1*, pp 271.

(3) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(4) Jorgensen, W. L.; Tirado Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(5) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205−1218.

(6) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. *Annu. Rev. Phys. Chem.* **2011**, *62*, 41−64.

(7) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Boca Raton, FL, 2008.

(8) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *J. Phys.: Condens. Matter* **2004**, *16*, R481−R512.

(9) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730−3737.

(10) Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624−1636.

(11) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.

(12) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.

(13) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.

(14) Chu, J.-W.; Voth, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *103*, 13111−13116.

(15) Chu, J.-W.; Voth, G. A. *Biophys. J.* **2006**, *90*, 1572−1582.

(16) Lyman, E.; Pfaendtner, J.; Voth, G. A. *Biophys. J.* **2008**, *95*, 4183−4192.

(17) Case, D. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285−290.

(18) Ma, J. *Structure* **2005**, *13*, 373−380.

(19) Bahar, I.; Rader, A. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.

(20) Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. *Chem. Rev.* **2010**, *110*, 1463−1497.

(21) Dykeman, E. C.; Sankey, O. F. *J. Phys.: Condens. Matter* **2010**, *22*, 423202.

(22) Ma, J.; Karplus, M. *J. Mol. Biol.* **1997**, *274*, 114−131.

(23) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 1−7.

(24) Tama, F.; Sanejouand, Y. H. *Protein Eng.* **2001**, *14*, 1−16.

(25) Cui, Q.; Li, G.; Ma, J.; Karplus, M. *J. Mol. Biol.* **2004**, *340*, 345−372.

(26) Petrone, P.; Pande, V. S. *Biophys. J.* **2006**, *90*, 1583−1593.

(27) Halioglu, T.; Bahar, I.; Erman, B. *Phys. Rev. Lett.* **1997**, *79*, 3090−3094.

(28) Bahar, I.; Atilgan, A. R.; Erman, B. *Fold. Des.* **1997**, *2*, 173−181.

(29) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505−515.

(30) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.

(31) Kundu, S.; Melton, J. S.; Sorensen, D. C.; Phillips, G. N. *Biophys. J.* **2002**, *83*, 723−732.

(32) Sen, T. Z.; Feng, Y. P.; Garcia, J. V.; Kloczkowski, A.; Jernigan, R. L. *J. Chem. Theory Comput.* **2006**, *2*, 696−704.

(33) Van Wynsberghe, A. W.; Cui, Q. *Biophys. J.* **2005**, *89*, 2939−2949.

(34) Hinsen, K.; Kneller, G. R. *J. Chem. Phys.* **1999**, *111*, 10766−10769.

(35) Hinsen, K. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 417−429.

(36) Ming, D.; Wall, M. E. *Phys. Rev. Lett.* **2005**, *95*, 198103.

(37) Erman, B. *Biophys. J.* **2006**, *91*, 3589−3599.

(38) Kondrashov, D. A.; Cui, Q.; Phillips, G. N. *Biophys. J.* **2006**, *91*, 2760−2767.

(39) Moritsugu, K.; Smith, J. C. *Biophys. J.* **2007**, *93*, 3460−3469.

(40) Yang, L.; Song, G.; Jernigan, R. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12347−12352.

(41) Riccardi, D.; Cui, Q.; Phillips, G. N. *Biophys. J.* **2009**, *96*, 464−475.

(42) Romo, T. D.; Grossfield, A. *J. Chem. Theory Comput.* **2011**, *7*, 2464−2472.

(43) Leioatts, N.; Romo, T. D.; Grossfield, A. *J. Chem. Theory Comput.* **2012**, *8*, 2424−2434.

(44) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144−150.

(45) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192−198.

(46) Kim, M. K.; Jernigan, R. L.; Chirikjian, G. S. *Biophys. J.* **2002**, *83*, 1620−1630.

(47) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807−822.

(48) Chu, J.-W.; Voth, G. A. *Biophys. J.* **2007**, *93*, 3860−3871.

(49) Doruker, P.; Jernigan, R. L.; Bahar, I. *J. Comput. Chem.* **2002**, *23*, 119−127.

(50) Kurkcuoglu, O.; Jernigan, R. L.; Doruker, P. *Polymer* **2004**, *45*, 649−657.

(51) Kurkcuoglu, O.; Turgut, O. T.; Cansu, S.; Jernigan, R. L.; Doruker, P. *Biophys. J.* **2009**, *97*, 1178−1187.

(52) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(53) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Comput. Chem.* **2008**, *4*, 435−447.

(54) Brooks, B. R.; Janežič, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522−1542.

(55) Swaminathan, S.; Ichiye, T.; van Gunsteren, W. F.; Karplus, M. *Biochemistry* **1982**, *21*, 5230−5241.

(56) Levy, R. M.; Perahia, D.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 1346−1350.

(57) Colletier, J. P.; Laganowsky, A.; Landau, M.; Zhao, M.; Soriaga, A. B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M. R.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16938−16943.

(58) Zimmerman, K. *J. Comput. Chem.* **1991**, *12*, 310−319.

(59) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.

(60) Grubmüller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997−999.

(61) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568−1581.

(62) Lu, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. *Biophys. J.* **1998**, *75*, 662−671.

(63) Krammer, A.; Lu, H.; Isralewitz, B.; Schulten, K.; Vogel, V. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1351−1356.

(64) Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16742−16747.

(65) Miller, Y.; Ma, B.; Nussinov, R. *Chem. Rev.* **2010**, *110*, 4820−4838.

(66) Xu, Z.; Paparcone, R.; Buehler, M. J. *Biophys. J.* **2010**, *98*, 2053−2062.

(67) Yoon, G.; Kwak, J.; Kim, J. I.; Na, S.; Eom, K. *Adv. Funct. Mater.* **2011**, *21*, 3454−3463.

(68) McCammon, J. A.; Gelin, B.; Karplus, M.; Wolynes, P. G. *Nature* **1976**, *262*, 325−326.
(69) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *141*, 423−447.
(70) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 4995−4999.
(71) Mchaourab, H. S.; Oh, K. J.; Fang, C. J.; Hubbell, W. L. *Biochemistry* **1997**, *36*, 307−316.
(72) de Groot, B. L.; Hayward, S.; van Aalten, D. M. F.; Amadei, A.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1998**, *31*, 116−127.
(73) Weaver, L. H.; Matthews, B. W. *J. Mol. Biol.* **1987**, *193*, 189−199.
(74) Sinev, M. V.; Sineva, E. V.; Ittah, V.; Haas, E. *Biochemistry* **1996**, *35*, 6425−6437.
(75) Arora, K.; Brooks, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18496−18501.
(76) Pontiggia, F.; Zen, A.; Micheletti, C. *Biophys. J.* **2008**, *95*, 5901−5912.
(77) Feng, Y.; Yang, L.; Kloczkowski, A.; Jernigan, R. L. *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 551−558.
(78) Müller, C. W.; Schulz, G. E. *J. Mol. Biol.* **1992**, *224*, 159−177.
(79) Rayment, I.; Rypniewski, W. R.; Schmidt-Base, K.; Smith, R.; Tomchick, D. R.; Benning, D. R.; Winkelmann, D. A.; Wesenberg, G.; Holden, H. M. *Science* **1993**, *261*, 50−58.
(80) Rayment, I.; Holden, H. M.; Whittaker, M.; Yohn, C. B.; Lorenz, M.; Holmes, K. C.; Milligan, R. A. *Science* **1993**, *261*, 58−65.
(81) Lombardi, V.; Piazzesi, G.; Ferenczi, M. A.; Thirlwell, H.; Dobbie, H.; Irving, M. *Nature* **1995**, *374*, 553−555.
(82) Molloy, J. E.; Burns, J. E.; Sparrow, J. C.; Kendrick-Jones, J.; Treager, R. T.; White, D. C. *Nature* **1995**, *378*, 209−212.
(83) Li, Y.; Brown, J. H.; Reshetnikova, L.; Blazsek, A.; Farkas, L.; Nyitray, L.; Cohen, C. *Nature* **2003**, *424*, 341−345.
(84) Adamovic, I.; Mijailovich, S. M.; Karplus, M. *Biophys. J.* **2008**, *94*, 3779−3789.
(85) Hvidt, S.; Nestler, F. H.; Greaser, M. L.; Ferry, J. D. *Biochemistry* **1982**, *21*, 4064−4073.
(86) Howard, J. *Mechanics of Motor Proteins and the Cytoskeleton*; Sinauer Associates: Sunderland, MA, 2001.
(87) Qin, Z.; Kreplak, L.; Buehler, M. J. *PLoS ONE* **2009**, *4*, e7294.
(88) Qin, Z.; Kreplak, L.; Buehler, M. J. *Nanotechnology* **2009**, *20*, 425101.
(89) Liu, B.; Huang, Y.; Jiang, H.; Qu, S.; Hwang, K. C. *Comput. Methods Appl. Mech. Eng.* **2004**, *193*, 1849−1864.
(90) Wackerfuβ, J. *Int. J. Numer. Meth. Eng.* **2009**, *77*, 969−997.