JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Resolution-Adapted All-Atomic and Coarse-Grained Model for Biomolecular Simulations

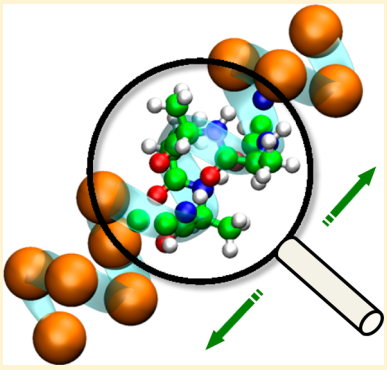
Lin Shen[†] and Hao Hu*[,†,‡]

[†]Department of Chemistry, University of Hong Kong, Pokfulam Road, Hong Kong, China

[‡]The University of Hong Kong-Shenzhen Institute of Research and Innovation, Kejizhong Second Road, Shenzhen, China

Supporting Information

ABSTRACT: We develop here an adaptive multiresolution method for the simulation of complex heterogeneous systems such as the protein molecules. The target molecular system is described with the atomistic structure while maintaining concurrently a mapping to the coarse-grained models. The theoretical model, or force field, used to describe the interactions between two sites is automatically adjusted in the simulation processes according to the interaction distance/strength. Therefore, all-atomic, coarse-grained, or mixed all-atomic and coarse-grained models would be used together to describe the interactions between a group of atoms and its surroundings. Because the choice of theory is made on the force field level while the sampling is always carried out in the atomic space, the new adaptive method preserves naturally the atomic structure and thermodynamic properties of the entire system throughout the simulation processes. The new method will be very useful in many biomolecular simulations where atomistic details are critically needed.



INTRODUCTION

Many important biological and chemical processes occur in complex heterogeneous environments. Computer simulation of these processes faces enormous difficulties, especially in selecting an accurate interaction Hamiltonian and achieving convergent phase space sampling. Although an ideal solution to satisfy both requirements is currently lacking, development and application of coarse-grained (CG) methods attracted much attention in recent years because of their excellent computational efficiency. The high computational efficiency of CG methods is attained by reducing the number of degrees of freedom and smoothing the effective (free) energy surface of the target processes as compared to atomistic models.[1−8]

Although the CG methods have demonstrated their effectiveness in the modeling of numerous complex processes,[9−18] many applications do require the availability of fine-grained atomistic structural information of the target system for practical chemical or biological reasons.[6,19] For example, atomic structure is absolutely needed if one plans to design new drug molecules on the basis of the simulation results. In this situation, accurate specifics of atomic interactions, not only for the binding pocket but also for other coupled allosteric sites of other protein molecules, are the key for the successful design of new ligand molecule. To take the advantage of the high computational efficiency of CG models, methods have been developed to recover or regenerate posteriorly the atomic structures from CG simulations, but the processes often suffer from the issue of nonunique mapping from CG to all-atomic (AA) models.[20,21]

Developing computational methods to combine CG and AA models and especially to preserve natively atomic structures of the target system become an active and important research topic. Several new methods have been developed along this line in recent years. The combination of AA and CG models can be spatially realized in either a homogeneous or a heterogeneous style. In the class of homogeneous mixing methods,[22−26] the system maintains its atomic structure, but the interactions are described simultaneously with both AA and CG models. Usually the two levels of model interactions are mixed together with a fixed ratio between the strengths. The mixing ratio is applied uniformly to the entire molecular system such that the identity of every atom in the system covers both AA and CG resolutions. Because different mixing ratios can lead to very different sampling efficiency and accuracy, multicanonical ensemble methods were often employed to allow the system interactions smoothly vary between AA Hamiltonian and CG Hamiltonian. In the class of heterogeneous mixing schemes, particles inherently of different level of resolutions are mixed together to construct a simulation system. This class of methods can be further divided into "domain-based" and "particle-based" coupling methods, respectively. In the domain-based approach,[27−31] a small region in space is selected. All particles within this region are described with AA force fields, or in other words at the AA resolution. A large CG region is defined such that particles within this region are described with CG models. To correctly model the identity change of particles traveling between the two regions, a buffer region is set up so that particles within can gradually and smoothly alter their identities between two levels of model. In the particle-based

Received: September 18, 2013
Published: May 7, 2014

   dx.doi.org/10.1021/ct401029k | J. Chem. Theory Comput. 2014, 10, 2528−2536

coupling approach,[32−36] the identity of particles, either coarse-grained or fully atomistic, are selected in advance and then fixed during the simulation. This class of methods followed very much the spirit of the widely employed combined quantum mechanical and molecular mechanical (QM/MM) methods.[37−39] Thus, several known issues in common QM/MM methods, for example, lack of flexibility for the size of the QM subsystem, also exist in the mixed AA/CG approach. Moreover, unlike the QM/MM methods where the interactions between the QM and MM subsystems are well-defined in electronic and/or atomic degree of freedoms, the interactions between the AA and CG particles were rarely studied before and must therefore require additional efforts to determine.[35] This factor becomes the main issue that hinders broader application of the approach to heterogeneous systems in addition to uniform molecular liquids and soft matter systems.

There is actually another large class of hierarchical methods, exemplified by the fast-multipole approach,[40] that could be useful in the development of new combined methods. Inspired by the structural hierarchy of many chemical and biochemical molecular systems, in this paper, we develop a new "AA+CG" approach where the resolution of interactions between all particles is adaptively adjusted in the simulation process. The key techniques are that the choice of theory is automatically adjusted at the force field level while the sampling is always carried out in the atomic space. Therefore, the AA+CG method takes the advantages of both the homogeneous and heterogeneous AA/CG mixing approaches. The atomic structure and thermodynamic properties of the entire system are preserved throughout the simulation processes, which consequently eliminate the need for an additional and more-or-less arbitrary mapping procedure to recover atomic structures after the simulations. The current method also integrates the efficiency of CG models and the accuracy of AA models as an important missing link in the broad spectrum of theoretical methods for the simulation of complex molecular processes from full quantum mechanics via atomistic molecular mechanics to pure CG models.

The rest of the paper is organized as follows. The theory is discussed first for the multiresolution model, the energy function, and the force for integration of atomic positions. The simulation details are provided in the next section, followed by the results. Discussion on the essence of this method and comparison to some other combined AA and CG methods is also provided.

THEORY

The essence of our present AA+CG method is that an all-atomic structural description of the target molecule is preserved throughout the simulation processes. A range-separated dual-resolution description for the interaction of the target system is employed as illustrated in Figure 1. The interactions between atoms at a short distance are described at an explicit AA level, while the interactions at a long distance are modeled purely at a CG level. At the intermediate distance, however, the AA and CG interactions will be mixed together in a distance-dependent fashion to allow smooth transition between two types of energy functions and forces. From the viewpoint of a selected atom, it interacts with the rest of the system in a heterogeneous way, starting from AA interactions with nearby groups and gradually evolving to fully CG forces with distant moieties. Since this distance-dependent heterogeneous interacting pattern is uniformly applied to all atoms in the simulation system, the development philosophy is clearly different from that of common QM/MM methods, but it is more similar to fragment-based linear scaling quantum chemistry methods.

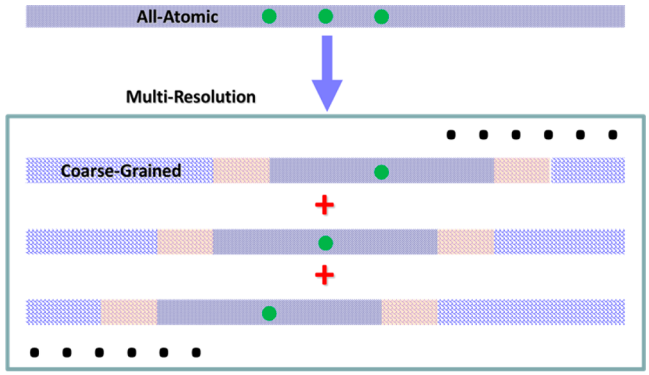


Figure 1. Scheme of the resolution-adapted all-atomic and coarse-grained model. For the interactions between a specific atom and the rest of the system, the current method calculates them adaptively using different models based on the interaction distances instead of a solely atomistic model.

We make one important assumption that there is a unique and nonoverlapping mapping between a specific set of atoms and a coarse-grained particle. This assumption is true for many currently developed CG force fields, especially those designed for protein molecules.[41,42] The position of a CG particle $\mathbf{R}_\alpha$ is determined by the center of mass of constituting atoms as

$$\mathbf{R}_\alpha = \sum_{i\in\alpha} m_i\mathbf{r}_i / \sum_{i\in\alpha} m_i \tag{1}$$

where m_i and $\mathbf{r}_i$ are respectively the mass and position of the atom i of a CG particle α. In this paper Latin letters i and j are used to represent atoms, while Greek letters α and β are used to label CG particles. The Hamiltonian of a molecular system of N atoms can be written as

$$H = \sum_i^N \frac{p_i^2}{2m_i} + V(\mathbf{r}, \mathbf{R}) \tag{2}$$

Here, p_i is the momentum of atom i. The first term of eq 2 is the kinetic energy described solely at the AA level. The second term is the potential energy modeled at the mixed AA+CG level. In general, the potential energy includes bonded covalent terms (i.e., bonds, angles, and dihedrals) and nonbonded terms (i.e., Coulombic and van der Waals (vdW) interactions). Since the bonded interactions are usually short-ranged, all of them will be described at the AA level. That is,

$$V^{\mathrm{cova}} = V_{\mathrm{AA}}^{\mathrm{cova}} \tag{3}$$

The form of the nonbonded interactions, on the other hand, will depend explicitly on the distance between CG particles. The nonbonded energy between two groups of atoms, each constituting a CG particle, is

$$V^{\mathrm{nonb}}(R_{\alpha\beta}) = [1 - \lambda(R_{\alpha\beta})]V_{\mathrm{CG}}^{\mathrm{nonb}}(R_{\alpha\beta}) + \lambda(R_{\alpha\beta}) \sum_{i\in\alpha, j\in\beta} V_{\mathrm{AA}}^{\mathrm{nonb}}(r_{ij}) \tag{4}$$

Here, r_{ij} is the distance between atoms i and j, and $R_{\alpha\beta}$ is the distance between CG particles α and β. Note that i must be smaller than j in the second term on the right-hand side if α and

dx.doi.org/10.1021/ct401029k | J. Chem. Theory Comput. 2014, 10, 2528−2536

β are the same CG particle. The coupling function $\lambda(R_{\alpha\beta})$ depends on the distance between CG particles α and β as

$$\lambda(R_{\alpha\beta}) = \begin{cases} 1 & R_{\alpha\beta} < R_1 \\ f(R_{\alpha\beta}, R_1, R_2) & R_1 \le R_{\alpha\beta} \le R_2 \\ 0 & R_{\alpha\beta} > R_2 \end{cases} \quad (5)$$

R_1 and R_2 are two cutoff distances for the transition between AA and CG models. The switching function $f(R_{\alpha\beta},R_1,R_2)$ takes the form as[43]

$$f(R_{\alpha\beta}, R_1, R_2) = 1 - 10\left(\frac{R_{\alpha\beta} - R_1}{R_2 - R_1}\right)^3 + 15\left(\frac{R_{\alpha\beta} - R_1}{R_2 - R_1}\right)^4 - 6\left(\frac{R_{\alpha\beta} - R_1}{R_2 - R_1}\right)^5 \quad (6)$$

When the distance between CG particles α and β is shorter than R_1, the atoms of α and β will interact with each other at the atomistic resolution, as the nonbonded interactions are fully described at the AA level. That is, $\lambda = 1$ and $V^{\text{nonb}} = V^{\text{nonb}}_{\text{AA}}$. When the distance is longer than R_2, atoms of each CG particle will not directly interact with each other; instead, they only experience a portion of the CG interactions. The nonbonded interactions are then described at a pure CG level; that is, $\lambda = 0$ and $V^{\text{nonb}} = V^{\text{nonb}}_{\text{CG}}$. At the intermediate distance ($R_1 \le R_{\alpha\beta} \le R_2$), the atoms of α and β interact with each other through both the AA and CG energy functions. In this case, the nonbonded interactions are calculated at both levels and mixed together with the weights determined from the coupling function $f(R_{\alpha\beta},R_1,R_2)$. The coupling function $f(R_{\alpha\beta},R_1,R_2)$ was selected to make smooth transition between 1 and 0. The function value and the first order derivative of the coupling function must be continuous at both boundaries of $R_{\alpha\beta} = R_1$ and $R_{\alpha\beta} = R_2$. We note that, in addition to the function currently employed in eq 6, the form of this coupling function and even the weights in eq 4 might be further refined according to the specific pair of AA and CG force fields used. However, this was not examined in the current work.

In our method the CG particles are treated as a virtual site by eq 1. Therefore, the CG forces are redistributed to the atomic degrees of freedom during the molecular dynamics (MD) integrations. As discussed, only the nonbonded CG forces need to be redistributed to atoms. The corresponding nonbonded force acting on atom i of a CG particle α can be derived from eq 4 as

$$\mathbf{F}_i^{\text{nonb}} = -\sum_{j\in\alpha, j>i} \frac{\partial V^{\text{nonb}}_{\text{AA}}(r_{ij})}{\partial \mathbf{r}_i} - \sum_{\beta\neq\alpha}\left\{\lambda(R_{\alpha\beta})\sum_{j\in\beta}\frac{\partial V^{\text{nonb}}_{\text{AA}}(r_{ij})}{\partial \mathbf{r}_i} + [1-\lambda(R_{\alpha\beta})]\frac{\partial V^{\text{nonb}}_{\text{CG}}(R_{\alpha\beta})}{\partial \mathbf{r}_i}\right\} - \sum_{\beta\neq\alpha}\left\{\frac{\partial\lambda(R_{\alpha\beta})}{\partial \mathbf{r}_i}\left[\sum_{j\in\beta} V^{\text{nonb}}_{\text{AA}}(r_{ij}) - V^{\text{nonb}}_{\text{CG}}(R_{\alpha\beta})\right]\right\} \quad (7)$$

where $-\partial V_{\text{AA}}/\partial \mathbf{r}_i$ is the force acting on atom i at the AA level, which is directly obtained from the AA potential function. However, the potential at CG level and the coupling parameter $\lambda(R_{\alpha\beta})$ are both explicit functions of the position of CG particle, $\mathbf{R}$, rather than atomic positions $\mathbf{r}$, so $\partial V_{\text{CG}}/\partial \mathbf{r}_i$ and $\partial\lambda/\partial \mathbf{r}_i$ in eq 7 are derived from eq 1 as

$$\frac{\partial V^{\text{nonb}}_{\text{CG}}(R_{\alpha\beta})}{\partial \mathbf{r}_i} = \frac{\partial V^{\text{nonb}}_{\text{CG}}(R_{\alpha\beta})}{\partial \mathbf{R}_\alpha}\frac{\partial \mathbf{R}_\alpha}{\partial \mathbf{r}_i} = \frac{m_i}{\sum_{i\in\alpha} m_i}\frac{\partial V^{\text{nonb}}_{\text{CG}}(R_{\alpha\beta})}{\partial \mathbf{R}_\alpha} \quad (8)$$

and

$$\frac{\partial\lambda(R_{\alpha\beta})}{\partial \mathbf{r}_i} = \frac{\partial\lambda(R_{\alpha\beta})}{\partial \mathbf{R}_\alpha}\frac{\partial \mathbf{R}_\alpha}{\partial \mathbf{r}_i} = \frac{m_i}{\sum_{i\in\alpha} m_i}\frac{\partial\lambda(R_{\alpha\beta})}{\partial \mathbf{R}_\alpha} \quad (9)$$

It should be noted that the derivative of $\lambda(R_{\alpha\beta})$ with respect to $\mathbf{R}$ is antisymmetric to the exchange of CG particle index α and β. Thus, all terms of eq 7 comply with Newton's third law everywhere. Clearly this mixed AA+CG model roots directly from the observation that group-based CG models can be used to describe the interactions between two sites when they are sufficiently far away. The further the distance between the two sites is, the coarser the interaction models.

The form of the CG force field is not specified up to this point. In principle, a closely chemistry-oriented CG force field like the UNRES force field[44−46] is preferred as it might be more compatible to the atomistic MM force fields. However, for the simplicity in implementation, MARTINI force field[32,45,46] is used in the current work. As will be discussed later, the new AA +CG method displayed minimal dependence on the CG force fields for many properties.

SIMULATION DETAILS

We outline the procedure of one MD integration step in the current method as follows:

(1) Update the CG configuration based on the atomic positions by eq 1.
(2) Calculate the AA+CG potential and force. The covalent contributions are evaluated at pure AA level as specified in eq 3. The nonbonded contributions are specified in eq 4 and 7, where the CG forces are distributed to the atomic degrees of freedom according to eq 8 and 9.
(3) Propagate the velocities and positions at the AA level forward in time using common (e.g., leapfrog)[47] integration scheme.

Compared with the AA resolution, the integration time step of the pure CG model is much larger (usually 15−30 times) because of the smoothed potential energy surface. Therefore, the multiple time step method[48] is used to split the total force into AA terms and CG terms depending on the model origin of the forces (i.e., AA or CG force fields).

Since the current approach will be most effective for the simulation of heterogeneous systems, we validated the adaptive-resolution method in the simulation of three protein systems, namely, the alanine-based α-helix peptide 3K(I) (AAAA-KAAAAKAAAAKA), the C-terminus β-hairpin of the B1 domain of protein G (residues 41−56, GEWTYD-DATKTFTVTE), and the ubiquitin protein (PDB entry code: 1UBQ).

Simulation setups were identical for all molecular systems. The interactions at the AA level were calculated using the CHARMM22/CMAP force field[49,50] combined with the generalized Born (GB) model.[51−53] The corresponding CG model was constructed with the standard mapping scheme used in the MARTINI model, which represents a protein residue with a backbone bead and several side-chain beads according to the residue type.[41,54,55] The transition region in which the AA and CG interactions were mixed together spanned from 6 to 9 Å (i.e., R_1 = 6 Å and R_2 = 9 Å in eq 5 and 6). The charged CG

particles interacted via the Coulombic energy function between point charges with a relative dielectric constant ϵ_{rel} = 15 to account for the screening effect. The vdW interactions at the CG level were described with the Lennard-Jones (LJ) potential, while the LJ parameters of the MARTINI force field were used with a slight modification: The parameters of particle type Nd, Na, and Nda were changed to that of the particle type N0 for the consistency with the implicit generalized Born model. The nonbonded interactions at the CG level vanish at the cutoff distance R_c = 12 Å. The GROMACS shift forms were used to remove the unwanted noise at the cutoff radius, shifting from 9 to 12 Å. The integration time step was 1 fs at the AA level and 10 fs at the CG level. Both the CG and AA nonbonded pairlists were updated every 20 fs. The temperature was maintained at 300 K using a Langevin thermostat with a friction coefficient of 2.0 ps^{-1}.[47]

To make comparison, both pure AA and CG simulations were carried out respectively and served as the references to the AA+CG simulations. In the reference AA simulations the CHARMM22/CMAP force field combined with GB model was used with a cutoff distance of 12 Å. To examine whether the CG interactions in the AA+CG method are important, a second set of simulations was carried out using a cutoff distance of 9 Å where the AA interactions die off in the AA+CG method. The integration time step was 1 fs. The temperature was also maintained at 300 K with a Langevin thermostat.

The MARTINI force field was used in the reference CG simulations. The vdW interactions were shifted from 9 to12 Å, while the electrostatic interactions were shifted from 0 to 12 Å.[41,55] As suggested,[55] the integration time step for the pure CG simulation was 20 fs, much larger than 10 fs used in AA +CG simulations. The temperature was maintained at 300 K using the Berendsen temperature coupling algorithm[56] with a relaxation time of 2.0 ps.

All the AA+CG simulations have been carried out for 100 ns by an in-house QM[4]D program package.[57] The pure CG simulations were performed with the GROMACS simulation package, version 3.3.1.[58] The secondary structure of each residue of these proteins was determined using the program DSSP.[59]

RESULTS

The root-mean-square deviations (RMSD) of the C_α atoms with respect to the native structure were computed and compared for simulations using the AA and AA+CG models. Table 1 summarizes the lowest, highest and mean values of the average RMSD of five trajectories of each molecular system. Figure 2 shows the RMSD as a function of simulation time for one representative trajectory of the three systems, respectively. The close agreements between the AA and AA+CG models in the distribution of C_α RMSD suggest that the resolution-adapted AA+CG scheme reproduced very well the structures sampled in AA simulations. We note that the bonded interactions in the MARTINI force field, the CG model used in our study, were originally introduced to maintain the secondary structure of proteins, which often overstabilize the protein native structure and underestimate structural fluctuations as compared to AA models. However, this trend was not observed in our simulations because only the nonbonded interactions at long distances were modeled at the CG level, which depend insignificantly on the secondary structures. Furthermore, the results of our method show minimal dependence on the selection of different nonbonded interactions in the MARTINI force field.

Maintaining proper secondary structural contents is an important criterion for the performance of protein simulations.[60−62] The population of secondary structures was analyzed for the simulations of three systems (Figure 3 and 4). It is obvious that the percentage of different structures in the current mixed AA+CG simulation is indeed close to those obtained at the AA level. The peptide 3K(I) adopted a stable α-helical structure between A3 and A14 (>90%) in both simulations (Figure 3a). In the C-terminus β-hairpin of GB1, both the two extended strands (W3T4Y5 and F12R14V14) and

Table 1. Lowest, Highest, and Mean Values of the Average RMSD (Å) from Five Trajectories of the Three Molecular Systems

		lowest	highest	mean
3K(I)	AA	1.17	1.39	1.26
	AA+CG	1.17	1.41	1.27
β-hairpin	AA	2.53	2.71	2.62
	AA+CG	2.55	2.79	2.69
ubiquitin	AA	2.32	2.96	2.70
	AA+CG	2.35	2.97	2.66

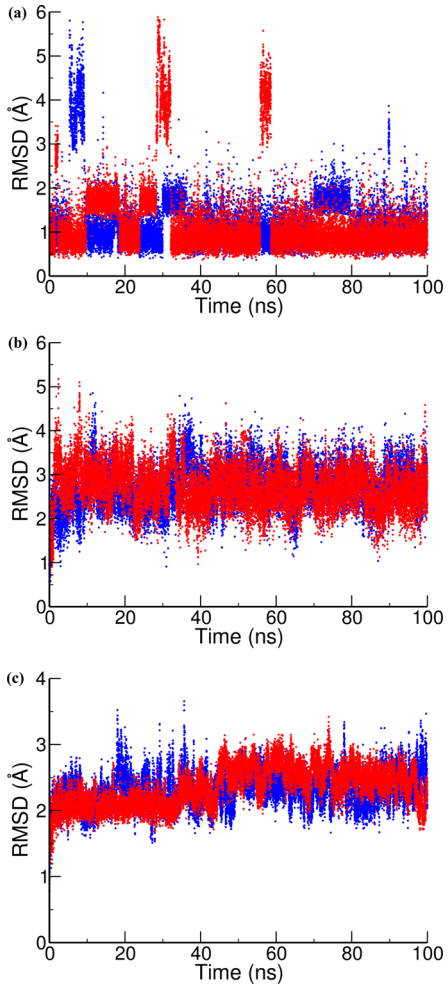


Figure 2. C_α RMSD as a function of time for one representative trajectory in three systems: (a) α-helix peptide 3K(I); (b) C-terminus β-hairpin of GB1; (c) ubiquitin. Different colors represent different interaction models (blue, AA; red, mixed AA+CG).

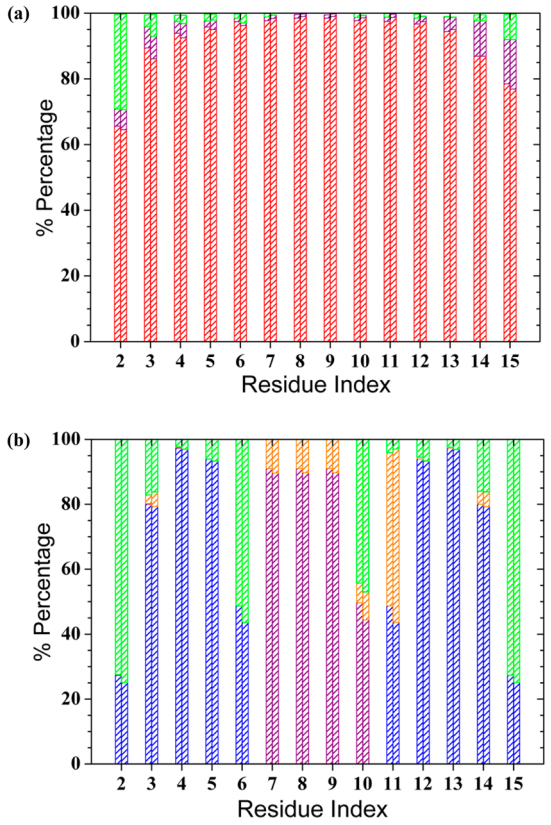


Figure 3. Percentage (%) of different secondary structures at each residue of (a) 3K(I) and (b) β-hairpin. For each residue, the bars at left side are from AA simulations and the right from AA+CG simulations. Colors represent different secondary structures (red, α-helix; violet, β-turn; green, coil; blue, extended; orange, β-bend).

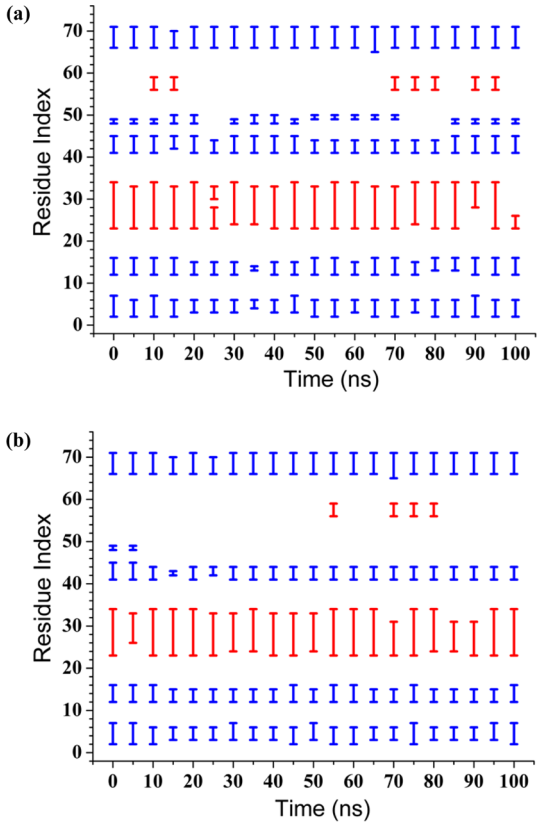


Figure 4. Secondary structure of ubiquitin as a function of simulation time for (a) AA and (b) mixed AA+CG simulations. Colors represent different secondary structures (red, α-helix; blue, extended).

the connecting β-turn maintain their structural characters. The results agree between two simulations (Figure 3b). Figure 4 shows that the structural elements of ubiquitin are also well preserved in both the AA and mixed AA+CG simulations. There might be a negligible difference for a short β-strand (residues 48−50) and a short α-helical fragment (residues 56−59) that was observed more frequently at the full AA level. Results of all three systems suggest that the new method can maintain both the α-helical and β-hairpin structures with performance comparable to the fully atomistic methods.

We also compared the distributions of structural parameters obtained from the AA and mixed AA+CG simulations (Figure 5). For the α-helix peptide 3K(I) (Figure 5a and b), distance distributions between residue pairs A6/A9, A6/A11, A4/A11, A3/A14, K5/K10, and A3/K10 all show excellent agreement between AA and AA+CG simulations. As these residue pairs span different regions in the resolution-adapted scheme, the good agreement suggests the correctness of our AA+CG model. This agreement is particularly important as one sees that the description of the interactions between some residue pairs will in fact change between the AA and CG models in the simulations. The distance distributions of the C-terminus β-hairpin of GB1 (Figure 5c) show almost identical results for residue pairs in the same strand (i.e., F12/V14 and T11/V14), between the AA and AA+CG simulations.

The structural results we report here seem to suggest that the AA force fields play the dominant role in determining the simulation structures. Two other results, however, suggested the importance of CG interactions in the current AA+CG method from different aspects. On one hand, pure AA simulation with a plain cutoff of 9 Å, mimicking the AA+CG model without the CG forces, produced larger structural RMSD (Supporting Information (SI) Figure S1). This suggests that the CG forces are important in our AA+CG model for stabilizing the native protein structure. Of course the degree of the importance of CG forces becomes an important issue to be examined in the AA+CG simulations with explicit solvent models. On the other hand, the results of radius of gyration (R_g) of ubiquitin also demonstrate the importance of CG in determining the molecular shape. In Figure 6, the distributions of R_g show obvious difference between three simulations with different models. The peak positions are 11.9, 11.6, and 11.6 Å, respectively for the AA, CG, and mixed AA+CG simulations. We attribute the difference between AA and AA+CG results to the nonbonded interactions in the CG force fields employed in the current simulations, as evidenced by the identical peak position between the AA+CG and pure CG simulations. Compared to the other two methods, broader distribution of R_g of pure CG simulations suggested larger fluctuation of the molecular shape in the simulations, presumably due to the lack of strong and often pair-specific electrostatic interactions in the CG nonbonded interactions. The smaller value of the peak position is likely caused by strong vdW packing interactions in the CG force field. The results of the AA+CG methods will be improved with the further development of the CG force fields, especially if the CG force fields are designed or refined with more consideration of the structural hierarchy of chemical systems. The comparison results reported here also suggest the our method could be very useful in refining CG force fields

dx.doi.org/10.1021/ct401029k | J. Chem. Theory Comput. 2014, 10, 2528−2536

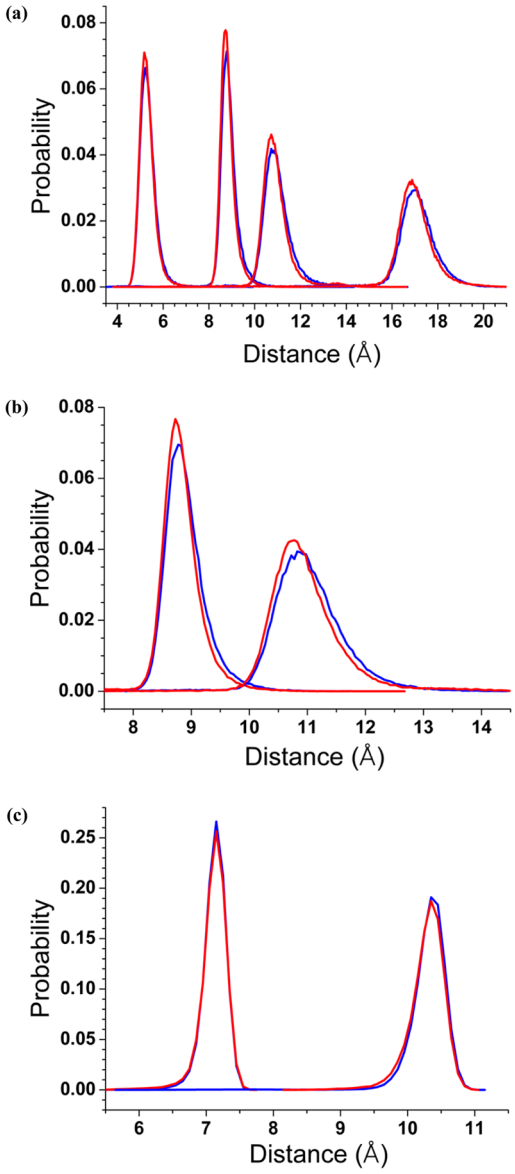


Figure 5. Distance distributions of 3K(I) and β-hairpin. From left to right, respectively: (a) A6/A9, A6/A11, A4/A11, and A3/A14 of 3K(I); (b) K5/K10 and A3/K10 of 3K(I); (c) F12/V14 and T11/V14 of β-hairpin. Different colors represent different levels (blue, AA; red, mixed AA+CG).

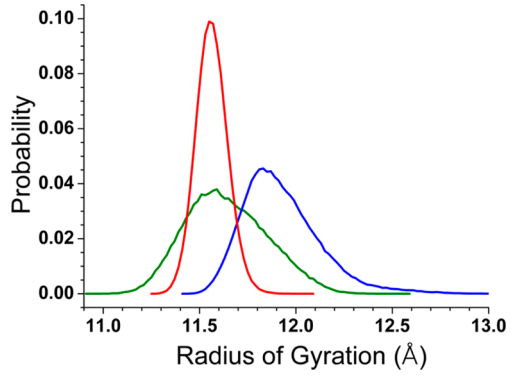


Figure 6. Distribution of radius of gyration of ubiquitin from simulations with different models (blue, AA; red, mixed AA+CG; green, CG).

with the coarse-graining approaches[63−65] to make it more consistent with AA force fields, in a way similar to defining MM force fields from QM methods. We also note that our AA or CG simulations results are in good agreement with previous work,[36] which suggested the validity of using them as the reference for the current AA+CG method.

DISCUSSION AND CONCLUSION

The motivation for the development of the current AA+CG method is to construct a systematic multiresolution model for the simulation of chemical systems, especially for the situation where atomic models are essential. Like the fragmentation technique made to the native full QM calculation of an entire molecule,[66,67] the current method can be naturally viewed as a fragment-based linear-scaling variant of the original AA methods. This comparison leads to two important consequences. First, the current method becomes an important link in the broad spectrum of hierarchical theoretical methods for the simulation of complex molecular systems.[68,69] In the order of increased efficiency and decreased accuracy, these methods include: native full QM method, fragment-based linear scaling QM methods including divide-and-conquer, QM/MM methods, pure MM methods, the current AA+CG method, AA/CG methods, and CG methods. Depending on the physics of the target process, one thus has the complete flexibility to choose appropriate simulation methods to reach a good balance between computational efficiency and accuracy. Second, systematic developments could be made to the current method to improve the accuracy on the basis of the hierarchy embedded in the theories.[69] Several improvements can be made to the current method. The first one is the choice of CG force field. MARTINI force field was used in a rather ad hoc way in the current work. As the design of the current method is based on the principle of hierarchical organization of physical interactions in molecular systems, a potentially better approach would be using a CG force field with more structural and chemical connections to the all-atom MM[44−46] or even QM force field. The transition between two levels of model was currently made on the criteria of interaction distance. An alternative and possibly better approach might be adjusted according to the magnitude of the forces between particles by using the multiscale coarse-graining methods.[70,71]

The current AA+CG method is the immediate approximation to the full atomistic method. Its computational cost, as compared to the atomistic models, is a critical factor that needs to be carefully analyzed. As shown in Table 2, the CPU time of the current method (183 s) is close to that of the AA simulation employing a (relatively small) plain cutoff of 9 Å (166 s). As the computational cost is approximately half of that of the reference AA simulation using a 12 Å cutoff, the AA+CG method would be efficient in simulations. Compared to AA models, the computational costs of CG force fields are small, especially when the multiple time step method was employed. The validity of using large time step for CG forces is justified by the NVE simulation results (SI Figure S2). The simulation using a time step of 10 fs for CG forces shows good conservation of energy while a time step of 20 fs shows a small energy drift (~0.5 kcal for 20 ps). Even though large time step was suggested for MARTINI force field before, considering the fact of small difference in time cost of two CG time steps reported in Table 2, we would recommend a time step of 10 fs for CG forces in the AA+CG scheme. We note that further refining and developing of computational algorithms are required if explicit

dx.doi.org/10.1021/ct401029k | J. Chem. Theory Comput. 2014, 10, 2528−2536

Table 2. CPU Time Spent for 10 000 Steps of MD Simulations of Ubiquitin with Different Interaction Models[a]

model	cutoff (Å)	AA time step (fs)	CG time step (fs)	pair lists updating frequency (fs)	time (s)
full AA/GB	12	1		10	315
full AA/GB	12	1		20	313
full AA/GB	9	1		20	166
AA+CG	12	1	1	20	213
AA+CG	12	1	10	20	183
AA+CG	12	1	20	20	181
AA+CG	12	1	20	200	179

[a]The AA and CG transition region was defined to be between 6 and 9 Å. The cutoff distance at CG level was 12 Å. The CPU used in this test is Intel Xeon E5620 (8 core, 2.4 GHz). The tests were carried out under strictly serial mode without using any special accelerating techniques.

solvent models, instead of implicit solvent models, are employed. In the current work with an implicit solvent model, the forces provided by CG force fields seem to be sufficient to supplement the short-range AA forces to stabilize the protein structure. Once explicit solvent models are employed, an immediate important question is how long-range electrostatic interactions are computed. Even though methods for computing long-range electrostatic interactions with AA force field have been well developed[72,73] and, in principle, they can be directly applied to CG force fields, how they could be efficiently applied to the mixed resolution methods such as the current AA+CG approach remains to be explored. The key question is whether the short-range electrostatic interactions computed with the AA model can be combined with long-range electrostatic interactions computed at the CG level—similar question could also be asked to the consistency of QM and MM force fields in the QM/MM methods. This question on long-range electrostatic interactions will be addressed in our following-up studies.

From the perspective of CG methods, one of the main goals for developing the current method bridging the AA and CG models is to maintain the atomistic structures. The preservation of full atomic degrees of freedom in the current method makes it naturally to preserve the thermodynamics of the target system. On this aspect, our method shares similar spirit with the methods that homogeneously mix AA and CG Hamiltonian.[22−26] The computational costs of the two methods should be comparable, which make both good candidates for resolution exchange simulations. But because of the difference in the spatial mixing scheme, the results of the two methods might show different degrees of dependence on the CG force fields.

Computational methods have been developed to recover atomistic structures from the CG models obtained in simulations.[20,21] Even though sharing with the same goal of obtaining atomistic structures, they differ significantly from the current method. These methods directly employed CG models in generating conformational ensembles. To revert back to atomistic models, an atomic structure is randomly generated subject to the constraint of geometry mapping between CG and AA models, and then relaxed through a simulated annealing process. Even though the sampling with CG models is efficient, the simulated annealing process is time-consuming—while this step is repeated for a large number of CG conformations. The greater challenge of this posterior processing method is the loss of associated ensemble weight during the recovering process: each original CG model possesses clear ensemble significance, but the recovered atomistic structure does not as the factor accounting for the contribution of phase space reduction between AA and CG models is generally unknown. Instead of using CG models, the current method directly samples the atomistic phase space. Even though then the sampling cost is relatively high compared to pure CG models, the elimination of additional expensive procedure to recover the AA structure would compensate its overall computational performance.

In conclusion, we have developed a new multiresolution method to simulate biomolecular systems at all-atomic and coarse-grained level. In our scheme, the resolution of interactions is determined by the distance between two particles rather than the position or number of one particle. It allows us to maintain an all-atomic structural description of the entire system during the simulations. Newton's third law is obeyed everywhere without compensation terms. The parameters in the AA and CG models can be used directly without the need to determine new parameters for the interactions between the particles at different resolutions. The simulations of three protein systems involving α-helical and β-hairpin structures obtained consistent results between the AA and the AA+CG models. We believe that this method is a valuable tool for biomolecular simulations.

■ ASSOCIATED CONTENT

Supporting Information

Figures for (1) the Cα RMSD of three systems simulated with GB model using different cutoffs and (2) NVE simulations using different time steps for the CG forces. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Email: haohu@hku.hk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the financial support from research grant council of Hong Kong, University of Hong Kong strategy research themes in Clean Energy and in Computation and Information, and National Science Foundation of China. Computational resources from the supercomputing facility in the University of Hong Kong are acknowledged.

■ REFERENCES

(1) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Boca Raton, FL, 2009.

(2) Darré, L.; Machado, M. R.; Pantano, S. Coarse-grained models of water. *WIRE Comput. Mol. Sci.* **2012**, *2*, 921.

(3) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. On developing coarse-grained models for biomolecular simulation: A review. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423.

(4) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801.

(5) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.

(6) Nielsen, S. O.; Bulo, R. E.; Moore, P. B.; Ensing, B. Recent progress in adaptive multiscale molecular dynamics simulations of soft matter. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12401.

(7) Baaden, M.; Marrink, S. J. Coarse-grain modelling of protein−protein interactions. *Curr. Opinion Struct. Biol.* **2013**, *23*, 878.

(8) Saunders, M. G.; Voth, G. A. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* **2013**, *42*, 73.
(9) Wang, Y.; Izvekov, S.; Yan, T.; Voth, G. A. Multiscale coarse-graining of ionic liquids. *J. Phys. Chem. B* **2005**, *110*, 3564.
(10) Becker, N. B.; Everaers, R. From rigid base pairs to semiflexible polymers: Coarse-graining DNA. *Phys. Rev. E* **2007**, *76*, 021923.
(11) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. DNA nanotweezers studied with a coarse-grained model of DNA. *Phys. Rev. Lett.* **2010**, *104*, 178101.
(12) Wolff, K.; Vendruscolo, M.; Porto, M. Coarse-grained model for protein folding based on structural profiles. *Phys. Rev. E* **2011**, *84*, 041934.
(13) Schulz, J. C. F.; Schmidt, L.; Best, R. B.; Dzubiella, J.; Netz, R. R. Peptide chain dynamics in light and heavy water: Zooming in on internal friction. *J. Am. Chem. Soc.* **2012**, *134*, 6273.
(14) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C. L.; Rodriguez-Ropero, F.; van der Vegt, N. F. A. Systematic coarse-graining methods for soft matter simulations—A review. *Soft Matter* **2013**, *9*, 2108.
(15) Derreumaux, P. Coarse-grained models for protein folding and aggregation. *Methods Mol. Biol.* **2013**, *924*, 585.
(16) Khalid, S.; Bond, P. J. Multiscale molecular dynamics simulations of membrane proteins. *Methods Mol. Biol.* **2013**, *924*, 635.
(17) Potoyan, D. A.; Savelyev, A.; Papoian, G. A. Recent successes in coarse-grained modeling of DNA. *WIRE Comput. Mol. Sci.* **2013**, *3*, 69.
(18) Daily, M. D.; Yu, H.; Phillips, G. N., Jr.; Cui, Q., Allosteric activation transitions in enzymes and biomolecular motors: Insights from atomistic and coarse-grained simulations. In *Dynamics in Enzyme Catalysis*; Klinman, J.; Hammes-Schiffer, S., Eds.; Springer: New York, 2013; Vol. *337*, pp 139.
(19) Zhou, H.-X. Theoretical frameworks for multiscale modeling and simulation. *Curr. Opinion Struct. Biol.* **2014**, *25*, 67.
(20) Rzepiela, A. J.; Schäfer, L. V.; Goga, N.; Risselada, H. J.; De Vries, A. H.; Marrink, S. J. Reconstruction of atomistic details from coarse-grained structures. *J. Comput. Chem.* **2010**, *31*, 1333.
(21) Stansfeld, P. J.; Sansom, M. S. P. From coarse grained to atomistic: A serial multiscale approach to membrane protein simulations. *J. Chem. Theory Comput.* **2011**, *7*, 1157.
(22) Christen, M.; Gunsteren, W. F. v. Multigraining: An algorithm for simultaneous fine-grained and coarse-grained simulation of molecular systems. *J. Chem. Phys.* **2006**, *124*, 154106.
(23) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. Resolution exchange simulation. *Phys. Rev. Lett.* **2006**, *96*, 028105 .
(24) Liu, P.; Shi, Q.; Lyman, E.; Voth, G. A. Reconstructing atomistic detail for coarse-grained models with resolution exchange. *J. Chem. Phys.* **2008**, *129*, 114103.
(25) Moritsugu, K.; Terada, T.; Kidera, A. Scalable free energy calculation of proteins via multiscale essential sampling. *J. Chem. Phys.* **2010**, *133*, 224105.
(26) Harada, R.; Kitao, A. Multi-scale free energy landscape calculation method by combination of coarse-grained and all-atom models. *Chem. Phys. Lett.* **2011**, *503*, 145.
(27) Praprotnik, M.; Site, L. D.; Kremer, K. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *J. Chem. Phys.* **2005**, *123*, 224106.
(28) Ensing, B.; Nielsen, S. O.; Moore, P. B.; Klein, M. L.; Parrinello, M. Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics. *J. Chem. Theory Comput.* **2007**, *3*, 1100.
(29) Izvekov, S.; Voth, G. A. Mixed resolution modeling of interactions in condensed-phase systems. *J. Chem. Theory Comput.* **2009**, *5*, 3232.
(30) Park, J. H.; Heyden, A. Solving the equations of motion for mixed atomistic and coarse-grained systems. *Mol. Sim.* **2009**, *35*, 962.
(31) Potestio, R.; Fritsch, S.; Español, P.; Delgado-Buscalioni, R.; Kremer, K.; Everaers, R.; Donadio, D. Hamiltonian adaptive resolution simulation for molecular liquids. *Phys. Rev. Lett.* **2013**, *110*, 108301.
(32) Rzepiela, A. J.; Louhivuori, M.; Peter, C.; Marrink, S. J. Hybrid simulations: Combining atomistic and coarse-grained force fields using virtual sites. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10437.
(33) Riniker, S.; Eichenberger, A.; van Gunsteren, W. Solvating atomic level fine-grained proteins in supra-molecular level coarse-grained water for molecular dynamics simulations. *Eur. Biophys. J.* **2012**, *41*, 647.
(34) Riniker, S.; Eichenberger, A. P.; van Gunsteren, W. F. Structural effects of an atomic-level layer of water molecules around proteins solvated in supra-molecular coarse-grained water. *J. Phys. Chem. B* **2012**, *116*, 8873.
(35) Riniker, S.; van Gunsteren, W. F. Mixing coarse-grained and fine-grained water in molecular dynamics simulations of a single system. *J. Chem. Phys.* **2012**, *137*, 044120.
(36) Wassenaar, T. A.; Ingólfsson, H. I.; Prieß, M.; Marrink, S. J.; Schäfer, L. V. Mixing MARTINI: electrostatic coupling in hybrid atomistic−coarse-grained biomolecular simulations. *J. Phys. Chem. B* **2013**, *117*, 3516.
(37) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic, and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227.
(38) Luzhkov, V.; Warshel, A. Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies. *J. Comput. Chem.* **1992**, *13*, 199.
(39) Florián, J.; Warshel, A. Langevin dipoles model for ab initio calculations of chemical processes in solution: Parametrization and application to hydration free energies of neutral and ionic solutes and conformational analysis in aqueous solution. *J. Phys. Chem. B* **1997**, *101*, 5583.
(40) Greengard, L.; Rokhlin, V. A fast algorithm for particle simulations. *J. Comput. Phys.* **1987**, *73*, 325.
(41) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819.
(42) Han, W.; Wu, Y.-D. Coarse-grained protein model coupled with a coarse-grained water model: Molecular dynamics study of polyalanine-based peptides. *J. Chem. Theory Comput.* **2007**, *3*, 2146.
(43) Leach, A. R. *Molecular Modelling: Principles and Applications*, 2nd ed.; Prentice Hall: Harlow, England, 2001.
(44) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.* **1997**, *18*, 849.
(45) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Ołdziej, S.; Scheraga, H. A. A united-residue force field for off-lattice protein-structure simulations. II. Parameterization of short-range interactions and determination of weights of energy terms by Z-score optimization. *J. Comput. Chem.* **1997**, *18*, 874.
(46) Liwo, A.; Kaźmierkiewicz, R.; Czaplewski, C.; Groth, M.; Ołdziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. United-residue force field for off-lattice protein-structure simulations: III. Origin of backbone hydrogen-bonding cooperativity in united-residue potentials. *J. Comput. Chem.* **1998**, *19*, 259.
(47) Van Gunsteren, W. F.; Berendsen, H. J. C. A leap-frog algorithm for stochastic dynamics. *Mol. Sim.* **1988**, *1*, 173.
(48) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1990.
(49) MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586.
(50) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400.

(51) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(52) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA continuum model for solvation: A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997**, *101*, 3005.
(53) Schaefer, M.; Bartels, C.; Karplus, M. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **1998**, *284*, 835.
(54) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2003**, *108*, 750.
(55) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812.
(56) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684.
(57) Hu, X. Q.; Hu, H.; Yang, W. T. *QM4D: An integrated and versatile quantum mechanical/molecular mechanical simulation package*. http://www.qm4d.info/ (accessed Sept 2013).
(58) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701.
(59) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577.
(60) Cino, E. A.; Choy, W. Y.; Karttunen, M. Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations. *J. Chem. Theory Comput.* **2012**, *8*, 2725.
(61) Patapati, K. K.; Glykos, N. M. Three force fields' views of the 3(10) helix. *Biophys. J.* **2011**, *101*, 1766.
(62) Zagrovic, B.; Sorin, E. J.; Pande, V. β-Hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **2001**, *313*, 151.
(63) Lyubartsev, A. P.; Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **1995**, *52*, 3730.
(64) Milano, G.; Müller-Plathe, F. Mapping atomistic simulations to mesoscopic models: A systematic coarse-graining procedure for vinyl polymer chains. *J. Phys. Chem. B* **2005**, *109*, 18609.
(65) Reith, D.; Putz, M.; Muller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624.
(66) Hu, X.; Jin, Y.; Zeng, X.; Hu, H.; Yang, W. Liquid water simulations with the density fragment interaction approach. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7700.
(67) Wu, A.; Xu, X. DCMB that combines divide-and-conquer and mixed-basis set methods for accurate geometry optimizations, total energies, and vibrational frequencies of large molecules. *J. Comput. Chem.* **2012**, *33*, 1421.
(68) Meier, K.; Choutko, A.; Dolenc, J.; Eichenberger, A. P.; Riniker, S.; van Gunsteren, W. F. Multi-resolution simulation of biomolecular systems: A review of methodological issues. *Angew. Chem., Int. Ed.* **2013**, *52*, 2820.
(69) Jin, Y.; Johnson, E. R.; Hu, X.; Yang, W.; Hu, H. Contributions of Pauli repulsions to the energetics and physical properties computed in QM/MM methods. *J. Comput. Chem.* **2013**, *34*, 2380.
(70) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *J. Chem. Phys.* **2008**, *128*, 244115.
(71) Wang, Y.; Noid, W. G.; Liu, P.; Voth, G. A. Effective force coarse-graining. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2002.
(72) Cisneros, G. A.; Karttunen, M.; Ren, P. Y.; Sagui, C. Classical electrostatics for biomolecular simulations. *Chem. Rev.* **2014**, *114*, 779.
(73) Sagui, C.; Darden, T. A. Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155.