An Approximate Method in Using Molecular Mechanics Simulations To Study Slow Protein Conformational Changes

Lijiang Yang and Yi Qin Gao*

Department of Chemistry, Texas A&M University, College Station, Texas 77843

Received: September 25, 2006; In Final Form: January 6, 2007

The broad range of characteristic motions in proteins has limited the applicability of molecular dynamics simulations in studying large-scale conformational transitions. We present an approximate method, making use of standard MD simulations and using a much larger integration time step, to obtain the structural changes for slow systematic motions of large complex systems. We show the applicability of this method by simulating the open to closed Calmodulin calcium binding domain conformational changes. Starting with the Ca^{2+}-bound X-ray structure, and after the removal of the Ca^{2+} ions, our calculation yielded intermediate conformations during the rearrangement of helices in each Ca^{2+} binding pocket, leading to a structure with a lowest rmsd of 1.56 Å compared to the NMR apo-calmodulin structure.

1. Introduction

Molecular dynamics (MD) simulation has been successfully applied to a large variety of molecular systems, especially biological systems. However, most of these systems, such as proteins, display a large variety of characteristic motions, ranging from the fast vibrations of covalent bonds to the slow and large-scale folding transitions. The fast motions set the upper bound for the integration time step of all-atom MD simulation to femtoseconds, and as a result, these MD simulations can only explore typically nanosecond time scales and their applicability is severely limited. To apply MD simulations to problems exhibiting a large range of time scales, its efficiency in obtaining dynamics of a longer time must be improved. Many methods have been designed to improve MD's efficiency in studying the slow processes. In one class of methods, the fast motions are separated and treated differently from the slower ones (e.g., simply being frozen, as in the programs SHAKE[1] and its variant RATTLE[2]), and as a result a larger integration time step is allowed. Or, by using multiple time steps,[3−7] the slow varying interactions are computed less frequently, such as in the RESPA method.[4,5] Alternately, to access events of longer time scales, one can speed up slow events so that they take place fast enough to be accessed by a MD simulation. A large number of methods fall into this latter category, for example, the simulated annealing method,[8,9] replica exchange method,[10−12] and the self-guided MD (SGMD).[13,14] In the simulated annealing and replica exchange methods, the system escapes from local energy minimums more effectively and thus different conformational regions can be accessed more easily. While in the SGMD, a guiding force estimated from an average of the instantaneous forces over a period of normal MD simulations is introduced to accelerate slow systematic motions. In this work, we introduce a different and approximate *nondynamical* method, which makes use of molecular dynamics simulations, with the hope to extend the approachable time scales in searching for intermediate conformations during the protein conformational change process. One similarity between this method and the earlier SGMD[13,14] is that average forces and velocities obtained from trajectories of normal MD simulations for a nonequilibrium spontaneous process are introduced into the equations of motions to facilitate the computation of the slow protein conformational change (and therefore, unfortunately, both methods are not capable of generating either actual dynamics or proper ensemble distributions). The large conformational changes that we are most interested in include those induced by local environment changes, such as ligand binding or release. One of the differences between the proposed method and SGMD is that the former, in addition to speeding up the slow motions, also slows down the fast motions, allowing a larger integration step size than that of SGMD. In the following, we first present the derivation and the algorithm of this method and illustrate how it works using harmonic oscillating systems. We then demonstrate the applicability of this method through studying the open to closed conformational changes of the calmodulin N-terminal and C-terminal domains.

2. Methods

As mentioned earlier, to capture the large-scale and slow protein conformational changes using all-atom simulations with the current computational power, one needs either to extend the simulation time accessed in the unit computing time or to speed up the slow systematic conformational motions. In the present approximate simulation method the two approaches are combined. A basic hypothesis in this method is that most of the short-range motions, such as those involving changes of bond lengths and angles, reach local equilibrium much faster than the large conformational changes, so that there exists a time scale t_1 during which only the fast but not the slow degrees of freedom have reached "pseudoequilibrium". Due to the readiness of equilibrating, the motions along fast degrees of freedom averaged over a period of time t_1 become much slower compared to that prior to averaging, while the direction and the speed of the slow and *directional* motions are much less affected by this averaging process. Therefore, when the instantaneous velocities are substituted by the average velocities (over a time period of t_1) to propagate motions, a large time step can be used in the calculations (for both fast and slow motions with the former

* Address correspondence to this author. E-mail: yiqin@mail.chem.tamu.edu.

10.1021/jp066289+ CCC: $37.00
Published on Web 02/24/2007

now slowed down). On the other hand, since the average velocities and forces obtained from a normal MD simulation are introduced into the equations of motions in the subsequent more approximate simulation, the trend of the slow motions exhibited in the normal MD can be amplified. Thus one can not only access a much larger time scale but also speed up the slow nonequilibrium processes. (As a result of the addition of the average velocities and forces, detailed balance is lost by the use of the present method.)

Assume that the system we are studying has two types of typical motions: fast motion, $\{x(t),\dot{x}(t)\}$,and slow motion, $\{y(t),\dot{y}(t)\}$. Although we are using x and y to denote the system, the derivation is general and can be applied to any other combination of motions. As a matter of fact, as seen below, the two types of motions can be along the same coordinate, e.g., a particle moves in a slowly drifting potential well. In the following we illustrate the method that allows us to study the slow motions in an approximate way. At first, a normal Newtonian dynamic simulation of N steps is carried out for the system using an integration time step of δt. An averaged velocity $\langle \dot{r}(t)\rangle_{t_1}$ ($r = x$ or y) is obtained from this trajectory of length t_1,

$$\langle \dot{r}(t)\rangle_{t_1} = \frac{r(t) - r(t - t_1)}{t_1} \tag{1}$$

where $t_1 = N\delta t$. Using the same trajectory, an effective force $g_r(t)$ is calculated:

$$g_r(t) = \frac{\langle \dot{r}(t)\rangle_{t_1} - \dot{r}(t - t_1)}{t_1} \tag{2}$$

The system is then propagated by using the following equation of motion, with the new position calculated as

$$r(t + \Delta t) = r(t) + \langle \dot{r}(t)\rangle_{t_1} \Delta t + \frac{1}{2}[\lambda_1 g_r(t) + \lambda_2 f_r(t)]\Delta t^2 \tag{3}$$

λ_1 and λ_2 are two constant parameters that will be determined below. Δt is the new integration time step size which is n times larger than the original time step size δt ($\Delta t = n\delta t$). $r(t + \Delta t)$ is then used in eq 1 to calculate $\langle \dot{r}(t + \Delta t)\rangle_{t_1}$ (notice that $r(t - t_1)$ in eq 1 becomes $r(t + \Delta t - t_1)$), and then $g_r(t + \Delta t)$ is calculated by eq 2, and so on and so forth). In eq 3 $r(t)$ is propagated by using the average velocity and the effective force, $\langle \dot{r}(t)\rangle_{t_1}$ and $g_r(t)$, in addition to the instantaneous force $f_r(t)$ on the coordinate r at time t. We show in the following how the application of eq 3 results differently for the slow and fast motions and why it provides an approximate description of the dynamics of the system. For the fast and near equilibrium motion along x, since t_1 is long enough, $\langle \dot{x}(t)\rangle_{t_1} \approx 0$, so that eq 3 becomes

$$\begin{aligned} x(t + \Delta t) &= x(t) + \frac{1}{2}\left[\lambda_1 \frac{-\dot{x}(t - t_1)}{t_1} + \lambda_2 f_x(t)\right]\Delta t^2 \\ &= x(t) + \left[-\lambda_1 \frac{n^2}{2N}\dot{x}(t - t_1) + \lambda_2 \frac{n^2}{2} f_x(t)\delta t\right]\delta t \end{aligned} \tag{4}$$

Letting $\lambda_1 = -2N/n^2$ and $\lambda_2 = 2/n^2$, eq 4 is simplified as

$$x(t + \Delta t) = x(t) + [\dot{x}(t - t_1) + f_x(t)\delta t]\delta t \tag{5}$$

One sees that the displacement calculated by eq 5 during the time interval Δt is the same as the displacement in the standard Newtonian dynamics simulation during a time interval of δt, with a velocity $\dot{x}(t - t_1)$. In other words, eq 5 resulted in a mapping of the displacement during each δt between 0 and t_1 to a displacement during a period of Δt between t_1 and Nt_1. Therefore by using eq 5, the motion of the fast degree of freedom has been slowed down by a factor of n ($=\Delta t/\delta t$). As a result, the distribution along x during the time interval (t_1, $(n+1)t_1$) is approximately the same as that shown during the standard Newtonian dynamics simulation (0, t_1), but also with a correction due to the instantaneous acceleration at time t.

On the other hand, since for the slow motion the variation of velocity is small during t_1, $\langle \dot{y}(t)\rangle_{t_1} \approx \dot{y}(t - t_1)$ and thus $g_y(t) \approx 0$, the application of eq 3 to the slow degree of freedom along y leads to

$$y(t + \Delta t) = y(t) + \langle \dot{y}(t)\rangle_{t_1} \Delta t + f_y(t)\delta t^2 \tag{6}$$

From eq 6, it is seen that the slow motion is propagated with an average speed under the influence of the instantaneous force, which is a function of both x and y. (It is partly for this reason that a correct distribution in x is needed.) One may interpret eq 6 as an extrapolation of the slow motion with a velocity of $\langle \dot{y}(t)\rangle_{t_1}$, and the instantaneous force f_y acts as the correction to the extrapolation. Because the use of an average velocity along the slow degrees of freedom (which is largely directional) tends to avoid the trapping of the system in the local minima during these largely downhill processes, the motion along the slow degrees of freedom becomes smoother and faster than that in the real dynamics (see the results on calmodulin, in particular, Figure 3). This effect is similar to that is seen when SGMD is used. We note here that since the same set of equations are applied to all degrees of freedom (in this example, x and y) the application of the current method does not require the separation of slow and fast modes. What is achieved by this method is an extrapolation of the motions along the slow degrees of freedom that exists but to a small extent in a normal MD simulation. In the mean time this extrapolation of the molecular motions has a minor influence on the equilibrium distribution along the fast degrees of freedom (e.g., bond stretching and bond bending). In addition, the method is more than a simple extrapolation, because the effective forces due to the molecular motions are also taken into account.

To summarize, the algorithm for the present simulation method is as follows:

(1) N steps of normal MD simulations are performed for a time length of t_1 with a step size of δt. During the normal MD simulation, coordinates and velocities of each atom in the molecular system are recorded at each n step.

(2) The average velocity is calculated as

$$\langle \dot{r}_i(t)\rangle_{t_1} = [r_i(t) - r_i(t - t_1)]/t_1 \tag{7}$$

from which an effective guiding force is evaluated:

$$g_i(t) = m_i \frac{\langle \dot{r}_i(t)\rangle_{t_1} - \dot{r}_i(t - t_1)}{t_1} \tag{8}$$

(3) Following the normal MD simulation and starting from time t, a larger time step $\Delta t = n\delta t$ is used to propagate the motions. And the acceleration at this time is calculated from the combination of the instantaneous force and the guiding force as

$$\tilde{\ddot{r}}_i(t) = \lambda_1 f_i(t)/m_i + \lambda_2 g_i(t)/m_i \tag{9}$$

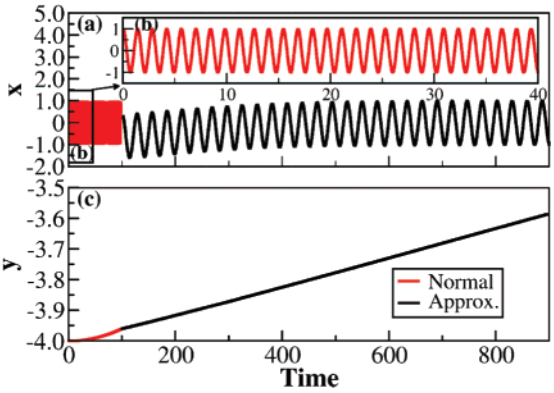

Figure 1. Application of the present approximate simulation method to a 2-dimensional harmonic oscillator: (a) 10 000 steps standard Newtonian dynamics simulation of the fast harmonic oscillation (red) and 4 000 steps approximate simulation for the same harmonic oscillation (black). (b) The first 4 000 steps of standard Newtonian dynamics simulation of the fast harmonic oscillation. (c) Slow harmonic oscillation: first, standard Newtonian dynamics simulation is run for 10 000 steps (red), and then the approximate simulation method is applied for another 10 000 steps (black, only 4 000 steps are shown here).

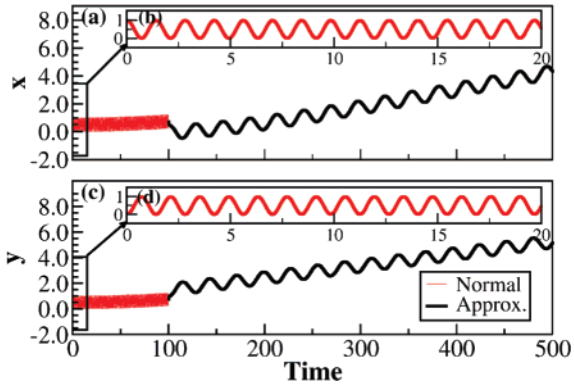

Figure 2. Application of the present approximate simulation method to a system with tightly coupled fast and slow harmonic oscillators: (a) 10 000 steps standard Newtonian dynamics simulation of the system along x (red) and 2 000 steps approximate simulation for the same system along x (black). (b) The first 2000 steps of standard Newtonian dynamics simulation of the system along x (red). (c) 10 000 steps standard Newtonian dynamics simulation of system along y (red) and 2 000 steps approximate simulation for the same system along y (black). (d) The first 2 000 steps of standard Newtonian dynamics simulation of the system along y (red).

where $\lambda_1 = -2N/n^2$ and $\lambda_2 = 2/n^2$, and the position is propagated as

$$r_i(t + \Delta t) = r_i(t) + \dot{r}_i(t)\Delta t + \tilde{\ddot{r}}_i(t)\Delta t^2/2 \qquad (10)$$

The new time step Δt and equations of motions (eqs 7−10) are used for N/n steps until all the positions and velocities recorded in step 1 have been used once.

(4) Repeat steps 1−3 until the end of the simulation.

3. Results

Harmonic Oscillator. To illustrate how the method works, we first applied the algorithm described above to a two-dimensional harmonic oscillator. In the example system, $\{x,\dot{x}\}$ is a fast motion with an elastic constant $k = 10$ N/m and $\{y,\dot{y}\}$ is a slow motion with an elastic constant $k = 10^{-6}$ N/m, so that the time scales of these two motions are well separated. First, a standard integration of motion was performed for 10 000 steps, with a step size of 0.01 s. The algorithm given above is then used with the step size set to 20 times larger, i.e., 0.2 s. (An even larger step size, i.e., 100 times, was also applied and showed in Figure S-1 in the Supporting Information.) Figure 1a,b shows the result for the fast vibration along x.

In Figure 1a the red solid line shows the first 10 000 steps of the standard integration of motion with the step size of 0.01 s. The coordinates and velocities during the 10 000 steps of the standard Newtonian dynamics are recorded every step. The black solid line in Figure 1a is the trajectory obtained by using the simulation method introduced in this study. Ten thousand integration steps were taken in the latter with a step size that is 20 times that used in the standard integration of motions to yield a trajectory of 2000 s. In Figure 1a only the first 4000 steps are shown. Motions after the 4000 steps of the approximate simulation repeat the oscillations between 600 and 900 s. As a comparison, the inserted Figure 1b shows the first 4000 steps of motions, using the standard integration of motions (note the 20 times difference between the units of time for parts a and b of Figure 1). Although the introduction of the average velocities and forces shifts a little the oscillating position of x during the initial stage, the oscillation recovers to its original position (around the origin) after a relatively short time (around 250 steps). The most noticeable difference between the trajectories shown in Figure 1b and the latter part of Figure 1a (after the introduction of average forces and velocities) is that the latter oscillates 20 times slower than the former. Figure 1c shows the result for the slow motion along y. Similarly, the red solid line represents the standard integration of motions and the black solid line represents the results calculated by using the approximate method. In contrast to the fast motion, the slow motion is not slowed down by the introduction of the average quantities. Instead, the motion propagates with the average velocity obtained from the standard integration of motions. The other two similar two-dimensional harmonic oscillators with different frequency separations ($\nu_{fast}/\nu_{slow} = 200$ and $\nu_{fast}/\nu_{slow} = 500$) were also tested and gave similar results. (See Figures S-2 and S-3 in the Supporting Information.) These results show that as long as the time scales of the fast and slow motions are well separated, so that an averaging time, which is (much) longer than the period of the fast motion and at the same time shorter than the slow motions of interest, can be chosen, the present simulation method selectively slows down the fast motion but not the slow motion. In the following we also apply the present approximate simulation method to a system composed of two tightly coupled harmonic oscillators: one is a fast oscillator with elastic constant $k_1 = 10$ N/m and the other is a slow oscillator with elastic constant $k_2 = 10^{-6}$ N/m. The system has the following potential energy: $U(x,y) = k_1(x - y)^2/2 + k_2(x - 10)^2/2$. The trajectory for the first 100 s was obtained from a standard integration of motions with a step size of 0.01 s and the rest was obtained by using eq 3. Figure 2a shows that the fast oscillation along the x-axis is slowed down and the global slow oscillation along the x-axis is speeded up by the use of eq 3. To compare with the original oscillation, the first 2000 steps of normal simulation are shown in Figure 2b. The oscillating frequency of the fast oscillation motion along x in approximate simulations is 20 times lower than that of normal simulation. For the motions on the y-axis, we have similar results. The fast local motion on the y-axis is slowed down while the slow global motion is speeded up (shown in Figure 2c,d). These results further validate the applicability of the present approximate simulation method to a realistic system in which almost all the motions are coupled together, and in the following, the

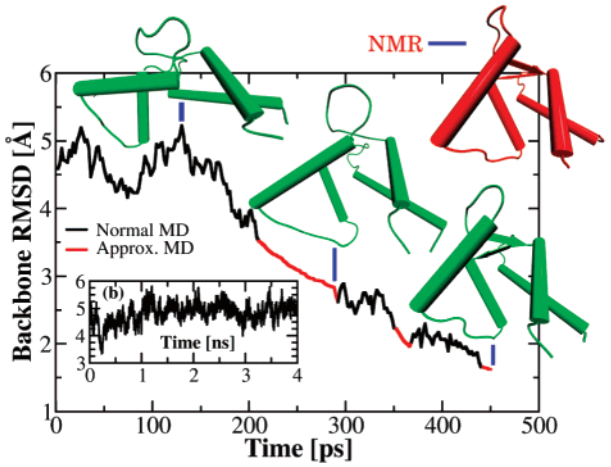

Figure 3. (a) Backbone rmsd between simulated structures and the NMR apo structure 1F70 for calmodulin N-terminal domain. Black denotes the normal simulation stage that is used to accumulate information for calculating the average. Red denotes the approximate simulation stage. (b) Backbone rmsd of a 4 ns normal MD simulation using the same initial structure as that used to obtain part a.

approximate simulation method introduced above is applied to more complex systems.

Calmodulin Calcium Binding Domain. In the following, we present results on the studies of the calmodulin conformational change in response to the release of Ca^{2+}. Calmodulin (CaM), an EF-hand calcium-binding and signal modulating protein, serves as a good example for theoretical studies of allosteric interactions and protein conformational changes in response to local environment changes because of its modest size (148 residues) and the abundance of experimental structures.[15−28] Both crystal Ca^{2+}-bound and NMR apo structures of CaM show a dumbbell-shaped molecule with two globular lobes, formed by the N- and C-terminal domains, respectively, separated by a flexible α-helix linker. The major conformational difference between the Ca^{2+}-bound and the apo CaM is that the Ca^{2+}-binding in the former leads to an exposure of the hydrophobic residues in the EF domain to the solution. These hydrophobic residues are buried inside the protein in apo CaM. Most of the previous computational studies on CaM focused on the equilibrium simulation of Ca^{2+}-bound CaM, apo CaM, CaM peptide complex, and the inherent flexibility of CaM.[29−40] Here we present simulations on conformational changes in the Ca^{2+}-binding domain due to the removal of Ca^{2+}, which is not observed in previous MD simulations.

The initial structure in our simulation is based on the high-resolution X-ray crystal structure of Ca^{2+}-bound CaM 1CLL,[18] but only the N- or C-terminal domain without Ca^{2+} was used and solvated in an octahedral TIP3P water box. The result on the CaM N-terminal domain (CaM-N), which has a better defined structure by NMR in the apo state,[41] is given first. The AMBER[42] ff99 force field was employed. Particle-mesh Ewald simulations under the NPT condition were conducted and SHAKE was employed to all the bonds involving hydrogen. The time step was set to 2 fs. The structure of CaM-N remains largely unchanged during a normal MD simulation of 4 ns and the rmsd fluctuates in the 3.5 to 6 Å range (Figure 3b). This remains true for a longer trajectory of 12 ns (data not shown). In the implementation of our simulation method, normal MD simulations were performed to generate trajectories (the states were saved every 40 fs) that were used for the propagation of motion using average forces and velocities. The time step used in the latter was 20 times larger than that in the normal MD,

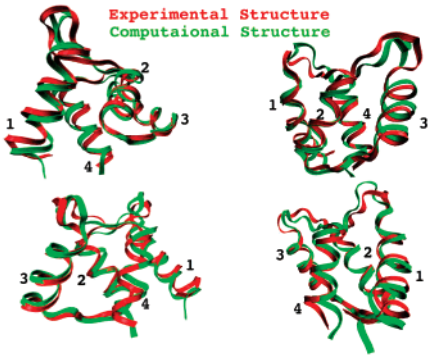

Figure 4. Alignment between the simulated structure (green) and the apo NMR structure 1F70 (red) for calmodulin N-terminal domain. (Four graphs from four different view ports are shown here.)

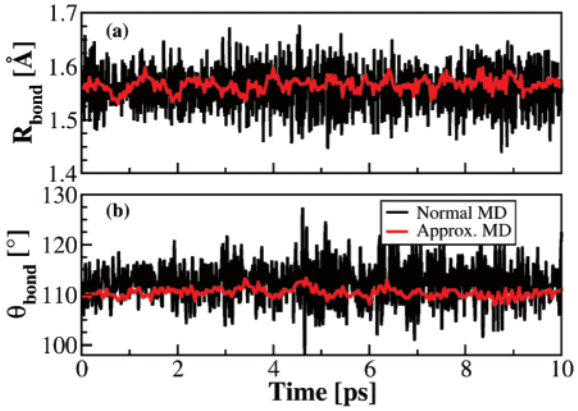

Figure 5. LEU18 (a) bond length (CB-CG) and (b) angle (CB-CG-CD1) comparison between normal MD simulation and the approximate simulation over 10 ps of each.

i.e., 40 fs. Two independent trajectories with similar results were obtained by using the current simulation method; one of them is shown in Figure 3a. In this figure, the black solid line segments are obtained from the normal MD simulation used to calculate average forces and velocities and the red solid line segments are for the approximate simulations. Three representative structures during the simulation are shown with three green cartoon representations and the NMR structure of apo CaM-N is shown in red. One can easily see that the protein undergoes a large conformational change during the simulation. Each pair of helices which make up the calcium binding site is rotated about 70° to transform from an open conformation to a closed conformation. The final structure obtained in the simulation is very similar to the experimental structure: the backbone rmsd is 1.56 Å, comparing to the NMR apo CaM-N structure 1F70.[19] The rmsd for all heavy atoms is 2.33 Å. Figure 4 shows the alignment of our final structure with the apo NMR structure 1F70 (to show each calcium binding site and each pair of helices clearly, four pictures each with a different view port are shown). The calculated and NMR experimental structure agree very well, as evident in the 1.56 Å rmsd. Further simulations along this trajectory do not improve the results. A normal MD simulation for the NMR apo CaM-N suggests that this rmsd of 1.56 Å is largely due to the intrinsic flexibility of the NMR structures (as shown in Figure S-4 in the Supporting Information, the backbone rmsd of this normal MD simulation for NMR apo CaM-N is around 1.5 Å). To examine the effect of eq 3 on the fast local motions, we show in Figure 5 the comparison of the fast local motions during the normal MD simulation and during the approximate simulation. In this figure, the black lines show the variations of bond length and angle during a normal MD

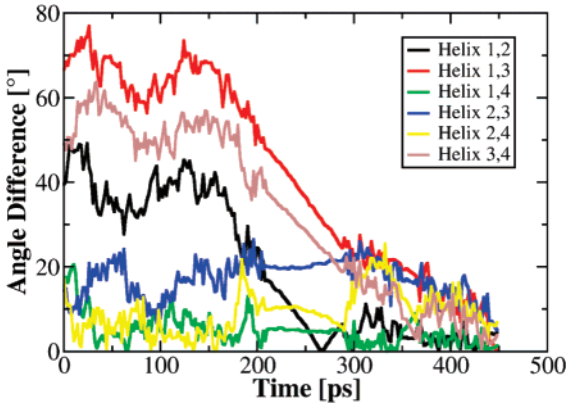

Figure 6. Interhelical angle differences between the present approximate simulation structure and the NMR apo structure 1F70. Angle differences between helix 1 (GLU3-PHE16) , helix 2 (THR26-LEU36), helix 3 (GLU42-ASP53), and helix 4 (PHE62-LYS72) are represented as follows: Helix 1,2 (black); Helix 1,3 (red); Helix 1,4 (green); Helix 2,3 (blue); Helix 2,4 (yellow); and Helix 3,4 (brown).

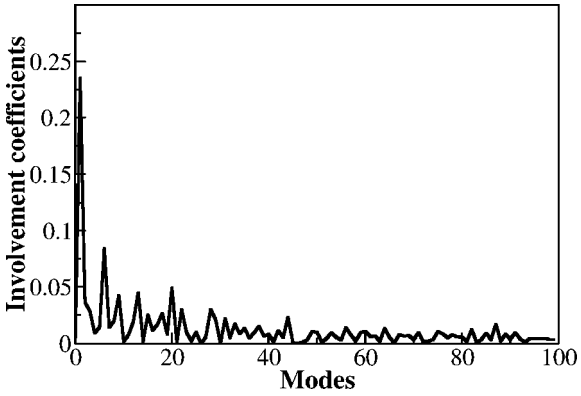

Figure 7. The involvement coefficients of the first 100 modes calculated from the quasiharmonic analysis of the 12 ns normal MD simulation trajectory of the calmodulin N-terminal domain with calcium ions removed.

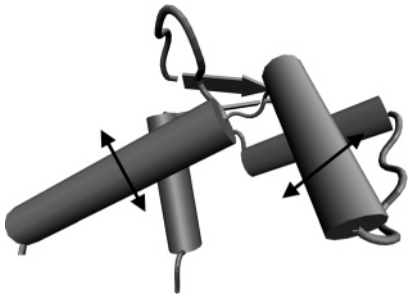

Figure 8. Motion representation of mode 2 calculated from the quasiharmonic analysis of the 12 ns normal MD simulation trajectory of the calmodulin N-terminal domain with calcium ions removed. The direction of the motion is showed by black arrows in the figure.

simulation while the red ones are for the approximate simulation. Apparently, in the approximate simulation, the variations of both bond lengths and angles are much smaller than those seen in the normal MD simulation, although in both simulations, the bond lengths and angles have similar values. These results show that when eq 3 is used, the fast local motions, such as the bond vibrations and angle bending, are dramatically slowed down, which then permits a much larger integration time step. As can be seen from the comparison between parts a and b of Figure 3, the approximate simulation method introduced here also facilitates a smooth protein conformational change, making it possible to capture large conformational changes in a relatively short computational time. One has to note, although, that due to the average velocities and effective forces added in the simulation process, this time itself does not have a physical meaning.

To show more details of how the conformational change occurs, the interhelical angles along with the conformational changes (from open state to closed state) are also calculated. The interhelical angles between each pair of helices 1−4 (Figure 4) were calculated during the simulation and absolute values of differences between the simulated and the experimental apo structure are shown in Figure 6. The initial open structure has large interhelical angle differences for $\angle\alpha_1,\alpha_2$, $\angle\alpha_3,\alpha_4$, and $\angle\alpha_1,\alpha_3$ compared to the closed conformation. During our simulations, all the interhelical angle differences, especially 1−2, 3−4, and 1−3, which are very different in the open and closed structures and serve as good indicators for the formation of the two Ca^{2+} binding sites, gradually decreased to 0°, almost in parallel to each other and to the total rmsd, indicating that the conformational changes of CaM occur in a concerted fashion.

To compare the conformational change obtained by the above approximate simulation with the normal modes of the calmodulin, we also carried out a quasiharmonic analysis of the CaM N-terminal domain. Our quasiharmonic analysis was applied to C_α based on the 12 ns normal MD simulation trajectory mentioned before. The modes cover frequencies ranging from 5.18 to 160.09 cm^{-1}. In addition, the involvement coefficient **I**

$$\mathbf{I} = |\mathbf{A}^T\mathbf{d}| \tag{11}$$

is also calculated. In eq 11, **d** is the projection of each mode on the direction of a specific conformational change. If $\mathbf{R_1}$ and $\mathbf{R_2}$ are two distinct conformers of the protein (for example, $\mathbf{R_1}$ denotes Ca^{2+}-loaded open conformation and $\mathbf{R_2}$ denotes the apo closed conformation), **d** is defined as

$$\mathbf{d} = \frac{\mathbf{R_2} - \mathbf{R_1}}{|\mathbf{R_2} - \mathbf{R_1}|} \tag{12}$$

The involvement coefficients of the first 100 modes of the CaM N-terminal domain in the direction of the conformational transition are shown in Figure 7. From Figure 7, it is clearly seen that mode 2 has the largest involvement coefficient of 0.24. The characteristic motion of mode 2 shown in Figure 8 (two black arrows show the direction of the motion) is very similar to the systematic motion captured in the approximate simulations. This means that the systematic motion we captured in the approximate simulations is indeed the real trend of the system. In a normal MD simulation, the slow systematic motions are concealed behind all kinds of local motions and need very long computational time to appear. On the contrary, the present approximate simulation method can detect the trend of motions by averaging over a relatively short normal simulation trajectory and push the system into this trend by extrapolating the motions based on the average velocities, average forces obtained from previous normal simulations, and instantaneous forces which act as local corrections.

As mentioned earlier, we also applied the present method to calmodulin C-terminal domain. The open to closed conformational change due to the removal of Ca^{2+} was also captured. Although the result for calmodulin C-terminal domain is not as good as that from the simulation of calmodulin N-terminal domain probably because the apo structure of calmodulin C-terminal domain is less stable, as seen from its more poorly defined NMR structure, the comparison with experiment is

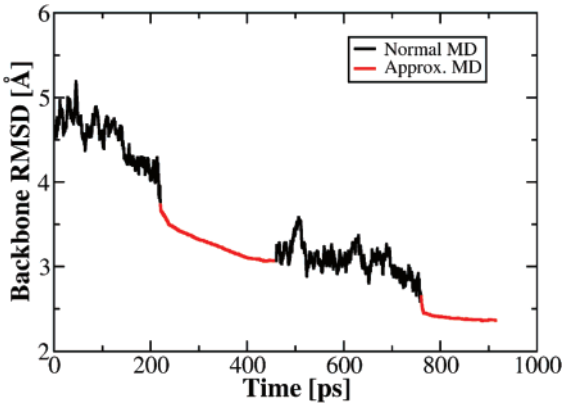

Figure 9. Backbone rmsd between simulated structures and the NMR apo structure 1F71 for calmodulin C-terminal domain. Black denotes the normal simulation stage that is used to accumulate information for calculating the average. Red denotes the approximate simulation stage.

Figure 10. Alignment between the simulated structure (green) and the apo NMR structure 1F71 (red) for calmodulin C-terminal domain.

satisfactory. The smallest backbone rmsd of the final calculated structure is 2.36 Å, compared to its NMR apo-structure 1F71.[19] The backbone rmsd during the present simulation trajectory can be seen in Figure 9 and the alignment of best simulated structure and NMR structure 1F71 is given in Figure 10.

4. Summary and Conclusions

An approximate method making use of MD simulations was developed to study the slow and large-scale protein conformational changes, in particular to search for intermediate conformations during protein conformational change processes which are too slow to be obtained by the standard molecular dynamic simulation method. The basic idea of the present method is to obtain an approximate distribution of the coordinates along the fast degrees of freedom using a relatively short normal dynamics simulation, and to subsequently use this distribution in the propagation of the motion by molecular mechanics simulations to capture the structure changes along the slow degrees of freedom. Through the usage of the present approximate simulation method, we not only make the use of a much larger time step size possible but also speed up the slow systematic motions (and therefore, unfortunately, lose the real dynamic information and canonical distribution). Thus the efficiency for the searching of intermediate conformations during the structural transitions is improved and it allows a quick search for viable transition conformations between two different known conformations, which are sometimes very difficult to obtain for complex systems. Studies on harmonic oscillating systems and on the conformational changes of the calmodulin N-terminal and C-terminal domains demonstrated the applicability and efficiency of the present method. Although application of this method breaks the detailed balance (as in the original form of SGMD), the intermediate conformations so obtained can be used to identify reaction coordinates of a complex system. When used in association with other methods, such as umbrella sampling, this method is expected to be helpful in obtaining thermodynamic as well as dynamic information.

Acknowledgment. We acknowledge ACS-PRF and the Camille and Henry Dreyfus Foundation for support. Y.Q.G. is a 2006 Searle Scholar.

Supporting Information Available: Application of the present approximate simulation method to a 2-dimensional harmonic oscillator with 100 times larger time step size than normal dynamic simulations; application of the present approximate simulation method to a 2-dimensional harmonic oscillator with 200 times frequency difference; application of the present approximate simulation method to a 2-dimensional harmonic oscillator with 500 times frequency difference; backbone rmsd of a 4 ns normal MD simulation of apo CaM-N; movies of the present approximate MD simulation trajectories for CaM-N and CaM-C. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(2) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24.
(3) Swindoll, R. D.; Haile, J. M. *J. Comput. Phys.* **1984**, *53*, 289.
(4) Tuckerman, M. E.; Martyna, G. J.; Berne, B. J. *J. Chem. Phys.* **1990**, *93*, 1287.
(5) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990.
(6) Izaguirre, J. A.; Reich, S.; Skeel, R. D. *J. Chem. Phys.* **1999**, *110*, 9853.
(7) Minary, P.; Tuckerman, M. E.; Martyna, G. *J. Phys. Rev. Lett.* **2004**, 93.
(8) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
(9) Wilson, C.; Doniach, S. *Proteins*: *Struct. Funct. Genet.* **1989**, *6*, 193.
(10) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96.
(11) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *J. Chem. Phys.* **2003**, *118*, 6664.
(12) Nagasima, T.; Sugita, Y.; Mitsutake, A.; Okamoto, Y. *Comput. Phys. Commun.* **2002**, *146*, 69.
(13) Wu, X. W.; Wang, S. M. *J. Phys. Chem. B* **1998**, *102*, 7238.
(14) Wu, X. W.; Wang, S. M. *J. Chem. Phys.* **1999**, *110*, 9401.
(15) Babini, E.; Bertini, I.; Capozzi, F.; Chirivino, E.; Luchinat, C. *Structure* **2006**, *14*, 1029.
(16) Babu, Y. S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1988**, *204*, 191.
(17) Bertini, I.; Del Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M.; Provenzani, A.; Zoroddu, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6841.
(18) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. *J. Mol. Biol.* **1992**, *228*, 1177.
(19) Chou, J. J.; Li, S. P.; Bax, A. *J. Biomol. NMR* **2000**, *18*, 217.
(20) Chou, J. J.; Li, S. P.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **2001**, *8*, 990.
(21) Finn, B. E.; Evenas, J.; Drakenberg, T.; Waltho, J. P.; Thulin, E.; Forsen, S. *Nat. Struct. Biol.* **1995**, *2*, 777.
(22) Houdusse, A.; Love, M. L.; Dominguez, R.; Grabarek, Z.; Cohen, C. *Structure* **1997**, *5*, 1695.
(23) Ishida, H.; Takahashi, K.; Nakashima, K.; Kumaki, Y.; Nakata, M.; Hikichi, K.; Yazawa, M. *Biochemistry* **2000**, *39*, 13660.
(24) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **1995**, *2*, 768.
(25) Rao, S. T.; Wu, S.; Satyshur, K. A.; Ling, K. Y.; Kung, C.; Sundaralingam, M. *Protein Sci.* **1993**, *2*, 436.

(26) Siedlecka, M.; Goch, G.; Ejchart, A.; Sticht, H.; Bierzynski, A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 903.
(27) Taylor, D. A.; Sack, J. S.; Maune, J. F.; Beckingham, K.; Quiocho, F. A. *J. Biol. Chem.* **1991**, *266*, 21375.
(28) Zhang, M.; Tanaka, T.; Ikura, M. *Nat. Struct. Biol.* **1995**, *2*, 758.
(29) Barton, N. P.; Verma, C. S.; Caves, L. S. A. *J. Phys. Chem. B* **2002**, *106*, 11036.
(30) Elezgaray, J.; Marcou, G.; Sanejouand, Y. H. *Phys. Rev. E* **2002**, 66.
(31) Fiorin, G.; Biekofsky, R. R.; Pastore, A.; Carloni, P. *Proteins*: *Struct. Funct. Bioinformat.* **2005**, *61*, 829.
(32) Komeiji, Y.; Ueno, Y.; Uebayasi, M. *FEBS Lett.* **2002**, *523*, 256.
(33) Likic, V. A.; Strehler, E. E.; Gooley, P. R. *Protein Sci.* **2003**, *12*, 2215.
(34) Shepherd, C. M.; Vogel, H. J. *Biophys. J.* **2004**, *87*, 780.
(35) VanderSpoel, D.; DeGroot, B. L.; Hayward, S.; Berendsen, H. J. C.; Vogel, H. J. *Protein Sci.* **1996**, *5*, 2044.
(36) Vorherr, T.; Kessler, O.; Mark, A.; Carafoli, E. *Eur. J. Biochem.* **1992**, *204*, 931.
(37) Wriggers, W.; Mehler, E.; Pitici, F.; Weinstein, H.; Schulten, K. *Biophys. J.* **1998**, *74*, 1622.
(38) Yang, C.; Kuczera, K. *J. Biomol. Struct. Dyn.* **2002**, *20*, 179.
(39) Yoon, J. H.; Jhon, M. S. *J. Mol. Struct.* **1993**, *295*, 193.
(40) Zuckerman, D. M. *J. Phys. Chem. B* **2004**, *108*, 5127.
(41) Nelson, M. R.; Chazin, W. J. *Protein Sci.* **1998**, *7*, 270.
(42) Case, D. A.; et al. *AMBER 8*; University of California: San Francisco, CA, 2004.