Binding free-energy calculation of an ion-peptide complex by constrained dynamics

Changjun Chen,[1] Yanzhao Huang,[1] Xuewei Jiang,[2] and Yi Xiao[1,*]

[1]*Biomolecular Physics and Modeling Group, Department of Physics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China*

[2]*School of Fashion, Wuhan Textile University, Wuhan 430073, Hubei, China*

(Received 31 December 2012; revised manuscript received 6 April 2013; published 14 June 2013)

Binding free energy is the most important physical parameter that describes the binding affinity of a receptor-ligand complex. Conventionally, it was obtained based on the thermodynamic cycle or alchemical reaction. These strategies have been widely used, but they would be problematic if the receptors and/or ligands have large conformational changes during the binding processes. In this paper, we present a way to calculate the binding free energy: constrained dynamics along a fragmental and high-dimensional transition path. This method directly considers unbound states in the simulation. The application to the calmodulin loop-calcium complexes shows that it is practical and the calculated relative binding affinities are in good agreement with experimental results.

DOI: 10.1103/PhysRevE.87.062705 PACS number(s): 87.15.ap, 87.15.hp

I. INTRODUCTION

Protein-ligand complexes play critical roles in biological processes. In the past, many different protein-ligand complexes have been studied deeply in theory and experiment. No matter what complexes we are studying, binding free energy is one of the important physical parameters we should deal with. The binding free energy is directly related to the relative stability between bound and unbound states of a complex or that of a complex with different ligands. Therefore how to calculate the binding free energy is an important mission in computational biology.

At present, the widely used method of calculating the binding free energy is based on the thermodynamic cycle [1,2]. This method first computes the solvation free energy from gas to solvent, ΔG_{solv}, for each monomer in the complex and the complex itself, by using the continuous solvent model [3,4]. Then it computes the interaction energy between monomers in the gas phase, $\Delta G_{\text{bind,gas}}$, by an all-atom molecular force field. Finally, the binding energy of the complex in solvent phase, $\Delta G_{\text{bind,solvent}}$, can be obtained by the thermodynamic cycle $\Delta G_{\text{bind,solvent}} = \Delta G_{\text{bind,gas}} - \Delta G_{\text{solv}}$. After considering the entropy contribution, the binding energy just turns into binding free energy naturally.

This classical method is efficient and easy to use. For example, Karplus and co-workers calculated the binding free energy of the insulin dimer by the thermodynamic cycle method [1] and the result is -11.9 kcal/mol, which is approximately close to the experimental value, -7.2 kcal/mol. Just with a similar method, Kollman's group carried out a calculation on the binding free energy of theophylline-RNA complex [2], and the result is -7.5 kcal/mol, comparable with the experimental value, -8.9 kcal/mol. These are successful applications of the thermodynamic cycle method.

However, in these calculations, it is difficult to consider the entropy contribution. For the molecule staying in the equilibrium state, based on the quasiharmonic approximation and normal mode analysis [5], the entropy could be divided into three terms: transitional, rotational, and vibrational terms. The last term is usually represented by the famous secondary-derivative *Hessian* matrix. Handling such a complicated matrix would be a time-consuming work. So in general, only a limited number of snapshots in the molecular dynamics (MD) trajectory are used in calculation. Moreover, when the molecule does not stay in the balanced structure during the binding, the quasiharmonic approximation breaks. In this case, entropy and enthalpy computation turns into a big problem.

On the other hand, some groups tried to do the binding free-energy calculation by an alchemical reaction, i.e., gradually shutting down the interactions between the receptor and ligand in the complex. This is not physical but easy for implementation. Until now, there were some successful applications of this kind of method [6–8], e.g., the binding free-energy calculation of *p*-xylene to the nonpolar cavity of the L99A mutant of the T4 lysozyme by Roux's group [7]. With the dual λ-REMD and *H*-REMD (Replica Exchange Molecular Dynamics) strategy, the calculated result, -4.93 kcal/mol, is rather close to the experimental one, -4.67 kcal/mol [7]. But theoretically, the standard alchemical method cannot consider the bound and unbound states simultaneously in the simulation. So the results will probably enforce or weaken the estimated binding affinity of the complex when the simulation is initiated from bound or unbound states [9,10].

Compared to alchemical reaction, another physics-based approach to calculate binding free energy is simulating the whole binding process. One end state in the process is that the receptor and ligand are separated far away from each other. The other end state is the native bound state. So in the simulation, there are no modifications on the force field. Using this approach, Woo and Roux performed the binding free-energy calculation between $p56^{lck}$ SH2 domain and a pYEEI peptide [11]. In the simulation, the peptide was pulled out of the receptor's binding pocket gradually, and the free-energy difference was computed by the free-energy perturbation method. The calculated binding free energy was in good agreement with experiment (error within 1 kcal/mol). The result also gave a downhill binding mechanism. However, using the conventional molecular dynamics, the receptor can only sample the conformations near the bound state. For the case that the receptor is rather flexible and has different

*Corresponding author: yxiao@mail.hust.edu.cn

conformations before and after binding ligand, this method may not be suitable. To sample the unbound state in the simulation, Laio and Parrinello used metadynamics [12] to study the binding of four typical complexes [13]. This method applied some additional Gaussian functions to the reaction coordinate to help the molecule escape from the free-energy minima. So the receptor could sample unbound state as well as bound state in the simulation. The simulated free-energy surfaces showed that there were some free-energy barriers on the binding path. Moreover, Kamiya's group performed flexible docking of the hen egg-white lysozyme and tri-NAG complex by the multicanonical molecular dynamics (McMD) [14]. The McMD simulation forced the molecule to sample evenly on the potential energy space that would allow the protein or ligand to have some probabilities to sample unbound states. The final free-energy profile also gave a high free-energy barrier along the reaction coordinate [15]. All these results indicate that the flexible binding processes are different from the rigid ones. It is necessary to consider the information from the unbound states in the calculation.

Recently, we implemented a method based on constrained dynamics, which integrates the free-energy differences along a predefined high-dimensional fragmental path [16]. This strategy can handle free-energy calculations for molecules in different metastable states, and shows better convergence than the conventional free-energy perturbation method. Here we extend it to the calculation of binding free energies between bound and unbound states of a complex. Although metadynamics [12] and McMD [14] can be used to sample unbound states in the simulation, they are different from our method. Both of them belong to the full-space sampling method. That is, to rebuild the free-energy profile, the molecules need to sample the whole space spanned by reaction coordinates (collective variables in metadynamics or potential energy in McMD). Such sampling could increase the simulation time. As a comparison, our simulation with constrained dynamics is a path-sampling method. The molecule is restricted to the transition path. Due to the limited sampling space, our method is more suitable for the simulations with a large number of collective variables.

The model in our study is the loop-calcium complex in calmodulin. Calmodulin can activate many proteins and enzymes, and take part in various biochemical reactions. Its role is strictly related to calcium concentration [17]. During the calcium-binding process, it changes from a closed structure to an open structure, exposes its hydrophobic cleft and is then ready to dock to and modulate other important proteins. This ion-induced dynamic behavior has been studied deeply for a long time [18], from ion-binding affinity [19–21], domain movement [22–25] to calcium-protein docking [26–31]. In this paper, we just pay attention to the ion-binding affinity of its four EF-hand loops. Some previous experiments revealed that the C-terminal domain has a stronger binding affinity (about tenfold) than the N-terminal domain in the intact protein [17,27,32], but the affinity of the N-terminal domain is higher when it is isolated [33,34]. This is interesting and it indicates that in the whole functional protein, these two domains are not assembled together in a simple way. The interdomain interactions may change the binding affinities of themselves [35]. To study the intrinsic binding ability of individual domains, Yang and co-workers grafted each EF-hand loop of the calmodulin onto a scaffold protein, domain 1 of CD2 [36], and found that the individual calcium affinities of the four loops follow the order I > II ≈ III > IV. This result again supports previous conclusions that the N-terminal domain shows higher calcium-binding affinity than the C-terminal domain in the isolated state.

Recently, Lepsik and Field studied this complex through simulation [37]. They calculated the calcium-binding free energies of the four EF-hand loops in the calmodulin by using a conventional thermodynamic cycle method. To improve accuracy in the energy computation, high-level density functional theory accompanied by molecular dynamics was used in the simulation. Their results show that the four loops favor the calcium in the order I > III > IV > II. It partially supports the experimental finding that loop I has the highest calcium-binding affinity while loop IV has the lowest binding affinity [36]; i.e., they found that loop II, instead of loop IV, has the lowest binding affinity. Due to the greatly different calcium affinities of loops I and II, the simulation results do not support that the isolated N-terminal domain has a higher binding affinity than the isolated C-terminal domain, which has been revealed in experiments [33,34].

In this paper, we use constrained dynamics to perform the binding free-energy calculation. The transition path connects the bound and unbound states in a high-dimensional space. After involving a virtual reaction coordinate, the constrained dynamics can give the free-energy profile of the binding process. The results show that the calculated relative binding free energies are in agreement with the experiment [36].

II. MATERIALS AND METHODS

A. Constrained simulation along a path composed of fragments

In the following, we will give a brief summary of the constrained dynamics. More information about its implementation can be found in our previous paper [16]. The constrained dynamics has been developed for many years [38–44]. One concise expression proposed by Schlitter and Klahn in 2003 is [45]

$$F_{0\to 1} = \int_0^1 \left\langle \frac{\partial H_c}{\partial \xi} \right\rangle_\xi d\xi - k_B T \ln\langle |Z|^{-1/2}\rangle|_0^1 = \int_0^1 \langle \lambda \rangle_\xi \, d\xi - k_B T \ln\langle |Z|^{-1/2}\rangle|_0^1. \quad (1)$$

Here ξ is a predefined reaction coordinate; 0 and 1 indicate the initial and final states respectively. H_c is the partition function of the constrained system; $\langle\rangle_\xi$ is the ensemble average of the constraint force at any ξ. λ is the *Lagrange*'s multiplier due to the constraint force, k_B is the Boltzmann factor, and T is temperature. The quantity $\boldsymbol{Z}$ is a L-dimensional matrix (L is the number of constraints in the simulation). Each element in the matrix is expressed by

$$Z_{\alpha\beta} = \sum_i \frac{1}{m_i} \frac{\partial \sigma_\alpha}{\partial x_i} \frac{\partial \sigma_\beta}{\partial x_i}. \quad (2)$$

In the equation, m_i is the mass of associated atom i and σ_α or σ_β is a constraint equation that corresponds to a different constraint.

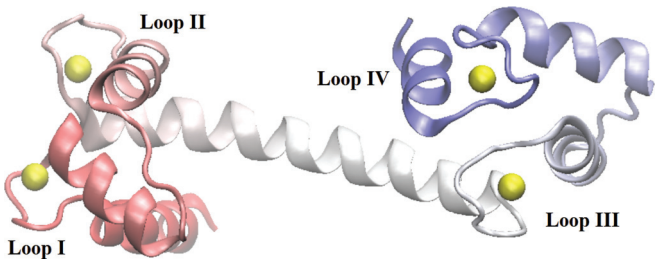

FIG. 1. (Color online) The crystal structure of calmodulin (PDBid: 1CLL). The yellow balls in the figure represent the calcium ions. The figure is produced by Visual Molecular Dynamics (VMD) [66].

Equation (1) and its equivalent form [46,47] in constrained dynamics have been proven successful in free-energy calculations, as when determining the potential of mean force over three-atom bending angles and four-atom torsion angles [47]. However, sometimes the necessary condition that ξ must be unique and differentiable with respect to the Cartesian coordinates of atoms is not satisfied.

To overcome this problem, we use one "virtual" reaction coordinate ξ to describe the transition path. $\xi = 0$ means initial state and $\xi = 1$ means final state. Here ξ increases monotonically, and corresponds to the transition process in a high-dimensional space with L degrees of freedom (or constraints). Similar to Ref. [48], for the kth snapshot in the path, its value ξ_k is defined as

$$\xi_k = \frac{\sum_{m=1}^{k} \sqrt{\sum_{i=1}^{L} \Delta\chi_{m,i}^2}}{\sum_{m=1}^{N} \sqrt{\sum_{i=1}^{L} \Delta\chi_{m,i}^2}}. \tag{3}$$

Here $\chi_{m,i}$ is the value of the constraint i in the mth snapshot. $\Delta\chi_{m,i}$ is the difference between $\chi_{m,i}$ and $\chi_{m,i+1}$. N is the total number of snapshots in the path.

For each fragment of the path (between successive snapshots), the free-energy difference can be expressed in a finite difference form [16]:

$$\Delta F(\xi_k) = \sum_{i=1}^{L} \langle\lambda_i\rangle_{\xi_k} \Delta\chi_i - k_B T \ln\langle|Z|^{-1/2}\rangle|_{\xi_k}^{\xi_{k+1}}. \tag{4}$$

Here $\langle\rangle_{\xi k}$ indicates that the system is constrained at $\xi = \xi_k$ in the simulation (corresponding to the kth snapshot), and λ_i is the Lagrange's multiplier for the ith constraint. Finally, the free-energy differences of all the fragments in the path are summed together to provide the complete free-energy profile for the transition.

B. Model

The model used in our work is calmodulin. It is a small protein with a highly conserved sequence in eukaryotic cells [17,22,49] (Fig. 1). It has a dumbbell-like structure, that is, two globular domains connected by a long and flexible helix. Each domain is constructed of two EF-hand motifs, which are well known in calcium-binding ability [50]. To check the binding affinity of each of the four EF-hand loops, we try to use constrained free-energy differences for them between their bound states and unbound states. Here both the bound state (PDBid: 1CLL) and unbound state (PDBid: 1CFD) of the calmodulin are taken from the protein data bank (PDB) database [51] (Fig. 2).

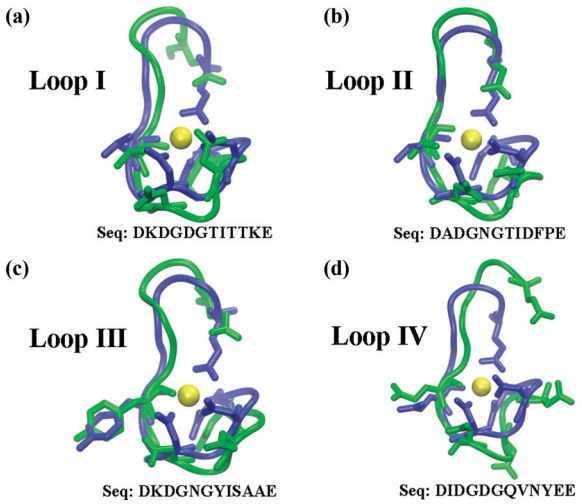

FIG. 2. (Color online) The structures of the four EF-hand loops of calmodulin in bound [blue (dark gray), PDBid: 1CLL] and unbound [green (light gray), PDBid: 1CFD] state. The yellow spheres in the figure represent the calcium ions. All the amino acids coordinating the calcium ions are illustrated in stick mode. The figure is produced by VMD [66].

C. Simulation detail

The optimization and MD software we use is TINKER [52] with the AMBER PARM96 force field [53]. But an additional free-energy calculation module is implemented by ourselves. The simulations are carried out at a normal temperature of 298 K, which is controlled by the Berendsen method [54]. The integration time step is 1.0 fs. To ensure the flexibility and mobility of the peptide, no bonds or angles are constrained. Moreover, there is no cutoff for both electrostatic and van der Waals interactions.

To simulate the aqueous environment, we use the Generalized Born Surface Area (GBSA) model [3,55] as an implicit solvent model. It is an approximation model of explicit solvent. Some previous studies have shown its reliability in protein folding [56–58] and in binding free-energy calculation [6]. In the GBSA model used in our work [3,55], the solvation free energy of calcium is -363.563 kcal/mol. It is approximately close to the simulation result (–360.300 kcal/mol) [59] and experimental result (–357.200 kcal/mol) [60] from other groups. The total simulation time depends on the number of intermediate structures in the path. This number is usually very large, generally over one hundred, even for short peptides. To increase the efficiency, we use the parallel strategy in the simulation, implemented by MPICH2 software [61]. In the simulation, according to the processor number, the total path is evenly divided into N parts (N is the number of processors), and each processor handles its own part. Finally, all the data from different processors are collected together to give the complete free-energy profile.

III. RESULTS AND DISCUSSIONS

In this work, a fragmental path-based constrained dynamic is applied to the binding free-energy calculation [16]. The predefined path directly connects the bound and unbound state of the loop-calcium complex. Here the bound states of the four EF-hand loop-calcium complexes (12 residues) are extracted from the crystal structure (PDBid: 1CLL), with proper caption atoms at two ends. And the corresponding unbound states are chosen from ion-free calmodulin structure (PDBid: 1CFD) (Fig. 2). All the structures are simulated for 1 ns in the solvent for equilibrium.

To do the formal calculation, an optimized path connecting the bound and unbound states is first built [16,62]. All the key dihedrals in the loop and the distance between the calcium and the loop (center of mass of coordinating atoms) are chosen as the constrained degrees of freedom. There are two stages in the transition path, named the binding stage and the approaching stage, respectively. During the first binding stage, all these key dihedrals are optimized from initial bound state to the final unbound state by restraint potential, and the calcium-loop distance is adjusted from that in the bound state to 20 Å. In this stage, the EF-hand loop begins to react to calcium and complete the whole unbound-to-bound structural transformation. For the second approaching stage, the calcium is perturbed from 20 to 200 Å away, with 0.1-Å intervals. During this stage, the EF-hand loop recognizes the distant calcium ion but stays in the unbound state in the process. The numerical free-energy result shows that this distance (200 Å) is far enough away to divide the complex into individual monomers, which is important for us to get the real binding free energy. Although this transition path is set by careful consideration, it may not be the real ion-binding process, but our object is to compute the complex binding free energy (i.e., binding affinity), which is a well-known state function. So any end-to-end path can satisfy this requirement.

In the following, we will give the details of the path to make sure it is really a transition path. First, the RMSD of all atoms in the four loops during the binding stage is shown in Fig. 3. The starting point of the path is the bound state, and the end point is the unbound state. The number along the x axis gives the index of the snapshot in the transition path. At the beginning, all the RMSDs are over 3.0 Å, which indicates the large structural difference between bound and unbound state. This calcium-loop binding case is definitely not a rigid docking process. So if only considering near-bound conformations in the calculation, as in the thermodynamic cycle method, it would give biased free-energy results. During the binding process, all the RMSDs decrease gradually, except that the RMSD of loop I increases for almost 1 Å, meaning that it undergoes an expanding-collapsing process. At the end of the path, all RMSDs fall down to 0.1 Å. It proves that all these four loops have reached the defined unbound states when the transition finished. This is a complete transition path.

Figure 4 gives the variance of calcium-loop distance along the path. The distance is defined by the position of calcium and the center of mass of all six coordinate atoms in EF-hand loops. All the loop complexes have the same set of coordinate atoms: one side chain oxygen atom (named "OD1") in residues 1, 3, 5;

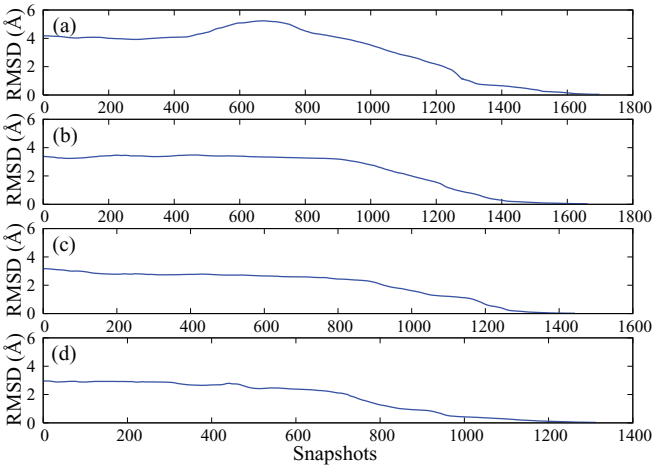

FIG. 3. (Color online) All-atom RMSD variance of the four EF-hand loops in the calcium-binding stage, from bound to unbound state. Here x axis indicates the index of snapshot in the transition path.

one backbone oxygen atom (named "O") in residue 7; and two side chain oxygen atoms (named "OE1" and "OE2") in residue 12 (shown in Fig. 2). From Fig. 4, we find that the calcium-loop distance varies, just as we expected in the first binding stage, from the value in the bound state to 20 Å. It validates the transition path again. Then in the second approaching stage, the distance is increased to 200 Å far away, 0.1 Å per snapshot.

Now, we will perform the formal free-energy calculation by constrained dynamics along the predefined bound-to-unbound transition path [with Eq. (4)]. For each calcium-loop complex, four independent trajectories are carried out. In the simulation, the CPU time spent on each individual snapshot is different. It depends on the error function defined as follows:

$$f = \sqrt{\sum_{i=1}^{L} \sigma(\lambda_{\chi_i})(\Delta\chi_i)^2 + (-k_B T)^2 \frac{\sigma(|Z|^{-1/2})}{\langle |Z|^{-1/2} \rangle_{\xi_k}}}. \quad (5)$$

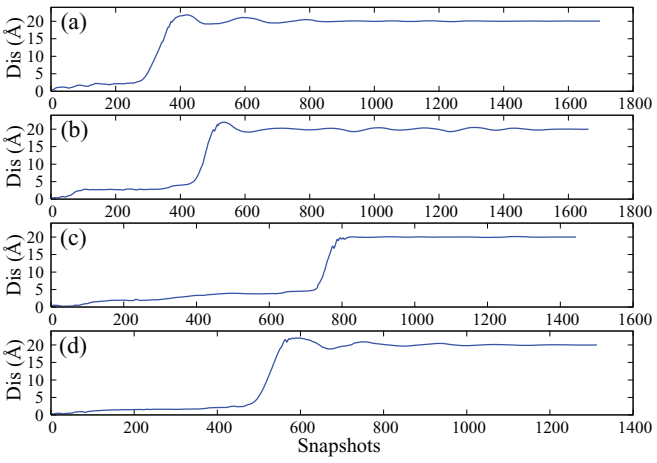

FIG. 4. (Color online) The variance of distance between calcium ion and EF-hand loops in the calcium-binding stage, from bound to unbound state. Here x axis indicates the index of snapshot in the transition path.

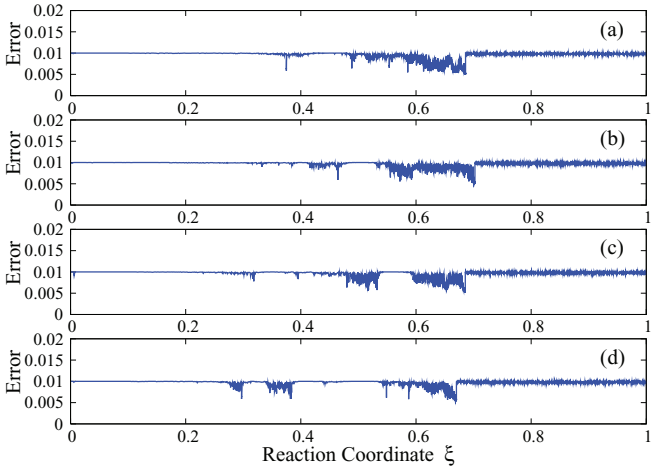

FIG. 5. (Color online) The iteration error via time for the four calcium-loop complexes (see the definition of the error function in the Materials and Methods section). All units are in kcal/mol.

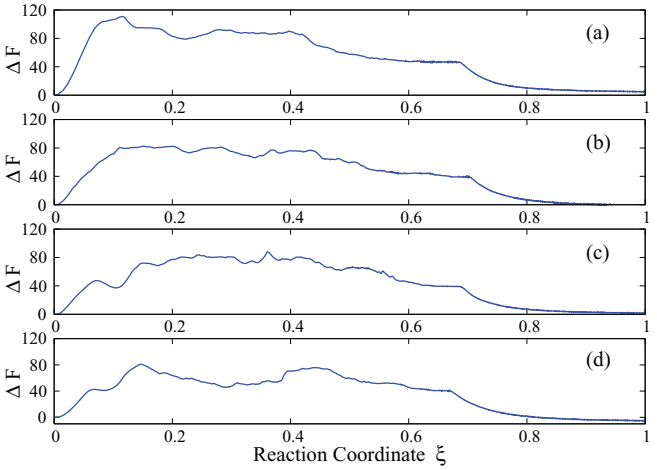

FIG. 6. (Color online) Free-energy variance along the transition path from bound state to unbound state. (a) Calcium-loop I complex; (b) calcium-loop II complex; (c) calcium-loop III complex; (d) calcium-loop IV complex. The calculation is performed by constrained dynamics. All the units are in kcal/mol.

Here $\sigma(x)$ is the square error of any variable x. For each snapshot in the path, the program will check the error on-the-fly after the first 10-ps relaxation period, and then the iteration will be terminated when the error is lower than 0.01 kcal/mol. With this strategy, the simulation will do more iteration steps automatically in the high-energy states. On the average, it takes about 90 ps for each snapshot in the path, and the total time for the whole path is about 300 ns.

As we noted before, the loop structure varies greatly in the binding stage (from unbound structure to bound structure), and the calcium ion moves greatly in the approaching stage (from 200 to 20 Å). So the total iteration error is the combination of the two sources. The final error function versus time is plotted in Fig. 5. It shows that the error is lower than 0.01 kcal/mol in the whole path and just satisfies the stop condition we set.

In the following, we will discuss the results of the free-energy calculation. The averaged free-energy value of all the four trajectories is shown in Fig. 6. Figures 6(a)–6(d) correspond to calcium-loop I–IV complexes. From the figure, it is clear that the free energy varies greatly during the whole transformation process. Initially, it increases to higher than 80 kcal/mol. This corresponds to the rotation of side chains in the loop. Then it goes down in two steps. The first step (decreasing to about 40 kcal/mol) is related to the movement of calcium ion, leaving from its docking pocket to 20 Å. The second decrease of free energy is related to the approaching stage in the transition path. As described above, in this stage, the loop stays in the unbound state and the calcium ion moves to 200 Å away. Although the free energies change greatly for all these four complexes, the final end-to-end free-energy differences are very small, which fall into the same order of magnitude of experimental results.

To further consider the flexibility of the molecule, we performed two more simulations in the bound and unbound state, respectively. This idea comes from Straatsma and McCammon's paper [63]. In the simulation, standard free-energy perturbation (FEP) with restraint potential was used. The free-energy difference was evaluated by

$$\Delta F = -k_B T \ln\langle \exp(-\Delta U)\rangle. \quad (6)$$

Here $\Delta U = k(\theta(\vec{r}) - \theta_0)^2$ is the restraint potential, with force constant $k = 10.0\,\text{kcal}/(\text{mol rad}^2)$. The θ includes the same key dihedrals that were used in the previous constrained dynamics. Both simulations last for 5 ns, and all the simulation conditions are the same as before. These two additional simulations can give the free-energy difference of the molecule between its rigid and flexible conformation. Combining these data with those from the transition path, we can have a complete free-energy difference or binding free energy.

Table I gives the final result. Here ΔF_{bound} and $\Delta F_{\text{unbound}}$ are the calculated free-energy differences at the bound and unbound state by the method of free-energy perturbation with restrained potential. ΔF_{path} is the calculated free-energy difference along the transition path by the method of constrained dynamics. The sum of all the differences above gives the total binding free energy ΔF for each calcium-loop complex between the bound and unbound states. Furthermore, we also compute the enthalpy difference ΔH (averaged potential energy) and entropy difference $-T\Delta S$.

From the table, it is clear that the free-energy difference between the bound and unbound states is relatively small, compared to the large variance of enthalpy and entropy. This is in agreement with the rule of entropy-enthalpy compensation [64], which has been observed in the previous free-energy calculation [1,2]. It indicates that the receptor-ligand binding in the biological system could be an enthalpy-favorable and entropy-unfavorable process. The loss of entropy is compensated by the decrease of enthalpy, so the variance of the total free energy is very small. It should be noted that the entropy contribution in our work is larger than that in Lepsik and Field's paper [37]. They calculated the entropy with the quasiharmonic approximation. That is, the molecule was assumed to be very stable in the simulation and its entropy could be divided into translational, rotational, and vibrational components. This approximation is reasonable for stable molecules. But it has been revealed in some papers that the calmodulin undergoes large structural movement in the unbound state ([24,25,65]),

TABLE I. Thermodynamic variable differences between bound and unbound state for each calcium-loop complex. The values in parentheses are statistical errors. Each one represents the square error of four independent trajectories. All units are in kcal/mol.

Complex	ΔF_{bound}[a]	ΔF_{path}[b]	$\Delta F_{\text{unbound}}$[c]	ΔF[d]	ΔH[e]	$-T\Delta S$[f]
Cal-loop I	4.432	−4.9	3.883	−5.4	−57.53	52.13
	(0.011)	(2.9)	(0.012)	(2.9)	(0.18)	(3.1)
Cal-loop II	4.057	0.7	3.395	0.0	−46.75	46.75
	(0.006)	(1.9)	(0.008)	(1.9)	(0.14)	(2.0)
Cal-loop III	3.972	−1.7	3.407	−2.3	−49.82	47.52
	(0.004)	(1.5)	(0.008)	(1.5)	(0.06)	(1.6)
Cal-loop IV	4.651	5.1	3.772	4.2	−62.0	66.2
	(0.013)	(0.8)	(0.005)	(0.8)	(0.4)	(1.2)

[a] ΔF_{bound}: The free-energy difference between restrained and unrestrained conformations in the bound state.
[b] ΔF_{path}: The free-energy difference along the transition path from unbound to bound state.
[c] $\Delta F_{\text{unbound}}$: The free-energy difference between restrained and unrestrained conformations in the unbound states.
[d] ΔF: The total free-energy difference (or binding free energy) between bound and unbound states.
[e] ΔH: The enthalpy difference between bound and unbound state, given by the averaged potential from all-atom force field.
[f] $-T\Delta S$: The entropy difference between bound and unbound state, obtained by $-T\Delta S = \Delta F - \Delta H$.

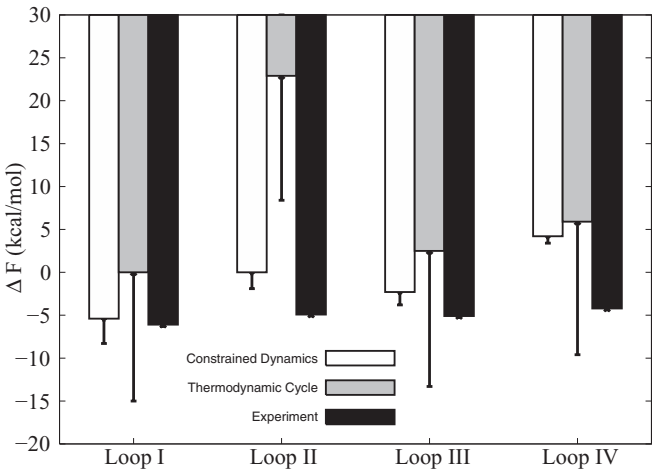

FIG. 7. Binding free energies of four calcium-loop complexes obtained by the constrained dynamics method (open boxes), together with data from thermodynamic cycle method (shaded boxes) [37] and experiment (solid boxes) [36] in other groups. Note that (1) The error bar in constrained dynamics is the square error of four independent trajectories (shown in Table I). (2) The original data based on thermodynamic cycle method are too large to be placed together with others, so they are shifted 224.5 kcal/mol up to make the value of the first complex zero.

so the quasiharmonic approximation may not be suitable for this system. As a comparison, in our work the entropy was calculated in a different way. First the binding free energy and the related enthalpy were obtained from the simulations. Then we used the formula $-T\Delta S = \Delta F - \Delta H$ to get the entropy component, so there was no quasiharmonic approximation in the calculation. We think this way is better for entropy calculation of flexible molecules.

Based on the data in Table I, loop I has the strongest calcium-binding affinity and loop IV the weakest. This difference comes from their special enthalpy and entropy contributions. For enthalpy, the averaged interaction energy between the calcium ion and four loops in the bound state (the minimized structure) is −276.146, −252.276, −270.182, −277.127 kcal/mol, respectively. This interaction includes van der Waals, electrostatic, and GB polarized energy. Among them, electrostatic interaction plays the critical role, especially for the fifth residue in the loops. The electrostatic energy between calcium ion and ASP5 in loop I or IV is about −200 kcal/mol, which is rather stronger than that between calcium and ASN5 in loop II or III. So the enthalpy contribution makes loops I and IV more favorable to the calcium ion than the other two loops. Moreover, Table I shows that, when the conformation transfers from unbound to bound state, the calcium-loop IV complex lost the largest amount of entropy. This disfavored contribution elevates the enthalpy difference and decreases the binding affinity of the fourth loop. But for loop I, the lost entropy is close to the other two loops, loops II and III. So as a result, calcium-loop I complex shows the lowest binding free energy.

Finally, we will show the unbound-to-bound free-energy differences for all four calcium-loop complexes, together with the data from the thermodynamic cycle method [37] and experiment [36] (Fig. 7). Here in the figure, the error bar of the data in our work is just the square error of four independent trajectories. In summary, the free-energy differences from calcium-loop I–IV complexes in our work are -5.4 ± 2.9, 0.0 ± 1.9, -2.3 ± 1.5, and 4.2 ± 0.8 kcal/mol, respectively. All the data have been illustrated in Fig. 7. The data from the thermodynamic cycle method are binding energy, not binding free energy. But the authors explained in their paper [37] that the variances of entropies between the four calcium-loop complexes were small. They omitted this term in the final results. So here we just assume their data as binding free energies. Since their data exceed the order of magnitude of the experiment, it is difficult to place them together with others. So for better illustration, all the data from the thermodynamic cycle method were shifted 224.5 kcal/mol up to make the value of the first complex as zero. From the figure, it could be found that the relative binding free energies of the four calcium-loop complexes in constrained dynamics are closer to experimental results. The ranked calcium-binding affinities of the four EF-hand loops are in the order, I > II ≈ III > IV. This is perfectly in agreement with experimental data, and validates the fact that the N-terminal domain has a smaller dissociation constant or higher calcium-binding affinity than the C-terminal domain [33,34].

IV. CONCLUSIONS

Protein-ligand complex is a typical molecular system in the living cell and the calculation of its binding free energy in a high level has always been important in computational biology. A reliable free-energy value can tell us the possibility of two monomers to bind to each other under certain condition

or which ligand among a large amount of candidates has the strongest binding affinity to the receptor.

Due to low price and high speed, the theoretical method is a potential way to cooperate with experiment. Up to now, the thermodynamic cycle method has been widely used in this field. It computes the absolute binding free energy by subtraction of other linkages in the thermodynamic cycle [1,2]. This method is efficient and practical, but it is not so easy to handle the flexible unbound state and the entropy contribution. The same challenge also exists in the simulation of alchemical reactions.

In this work, we use the constrained dynamics to calculate the absolute binding free energy of the complex, exactly from unbound to bound state. The final test on the calcium-loop complexes shows that this method is practical. Compared to experiment [36], it gives proper relative binding affinities. So this simulation method based on constrained dynamics may be an alternative and promising tool to study the protein-ligand binding problem.

ACKNOWLEDGMENTS

This work is supported partially by the NSFC under Grants No. 30800166, No. 11074084, and No. 11174093, and the Fundamental Research Funds for the Central Universities (Grant No. 2013QN018).

[1] V. Zoete, M. Meuwly, and M. Karplus, Proteins: Struct., Funct., Bioinf. **61**, 79 (2005).
[2] H. Gouda, I. D. Kuntz, D. A. Case, and P. A. Kollman, Biopolymers **68**, 16 (2003).
[3] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, J. Phys. Chem. A **101**, 3005 (1997).
[4] B. Honig, W. Rocchia, and E. Alexov, J. Phys. Chem. B **105**, 6507 (2001).
[5] B. R. Brooks, D. Janezic, and M. Karplus, J. Comput. Chem. **16**, 1522 (1995).
[6] M. Lapelosa, E. Gallicchio, and R. M. Levy, J. Chem. Theory Comput. **8**, 47 (2012).
[7] W. Jiang and B. Roux, J. Chem. Theory Comput. **6**, 2559 (2010).
[8] L. Wang, B. J. Berne, and R. A. Friesner, Proc. Natl. Acad. Sci. USA **109**, 1937 (2012).
[9] Y. Q. Deng and B. Roux, J. Chem. Theory Comput. **2**, 1255 (2006).
[10] D. L. Mobley, J. D. Chodera, and K. A. Dill, J Chem. Theory Comput. **3**, 1231 (2007).
[11] H. J. Woo and B. Roux, Proc. Natl. Acad. Sci. USA **102**, 6825 (2005).
[12] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. USA **99**, 12562 (2002).
[13] F. L. Gervasio, A. Laio, and M. Parrinello, J. Am. Chem. Soc. **127**, 2600 (2005).
[14] N. Nakajima, H. Nakamura, and A. Kidera, J. Phys. Chem. B **101**, 817 (1997).
[15] N. Kamiya, Y. Yonezawa, H. Nakamura, and J. Higo, Proteins: Struct., Funct., Bioinf. **70**, 41 (2008).
[16] C. Chen, Y. Huang, and Y. Xiao, Phys. Rev. E **86**, 031901 (2012).
[17] D. Chin and A. R. Means, Trends Cell Biol. **10**, 322 (2000).
[18] L. Konermann and J. X. Pan, Biochemistry **49**, 3477 (2010).
[19] H. Y. Park, S. A. Kim, J. Korlach, E. Rhoades, L. W. Kwok, W. R. Zipfel, M. N. Waxham, W. W. Webb, and L. Pollack, Proc. Natl. Acad. Sci. USA **105**, 542 (2008).
[20] J. Stigler and M. Rief, Proc Natl. Acad. Sci. USA **109**, 17814 (2012).
[21] M. Zhang, C. Abrams, L. P. Wang, A. Gizzi, L. He, R. Lin, Y. Chen, P. J. Loll, J. M. Pascal, and J. F. Zhang, Structure **20**, 911 (2012).
[22] G. Naray-Szabo, D. K. Menyhard, and G. M. Keseru, Curr. Comput.-Aided Drug Des. **5**, 264 (2009).
[23] K. Kuczera, C. Yang, and G. S. Jas, J. Biomol. Struct. Dyn. **19**, 247 (2001).
[24] W. Wriggers, E. Mehler, F. Pitici, H. Weinstein, and K. Schulten, Biophys. J. **74**, 1622 (1998).
[25] D. Vigil, S. C. Gallagher, J. Trewhella, and A. E. Garcia, Biophys. J. **80**, 2082 (2001).
[26] H. Kurokawa, M. Osawa, H. Kurihara, N. Katayama, H. Tokumitsu, M. B. Swindells, M. Kainosho, and M. Ikura, J. Mol. Biol. **312**, 59 (2001).
[27] S. W. Vetter and E. Leclerc, Eur. J. Biochem. **270**, 404 (2003).
[28] T. M. Lakowski, G. M. Lee, B. Lelj-Garolla, M. Okon, R. E. Reid, and L. P. McIntosh, Biochemistry **46**, 8525 (2007).
[29] T. Wyttenbach, M. Grabenauer, K. Thalassinos, J. H. Scrivens, and M. T. Bowers, J. Phys. Chem. B **114**, 437 (2010).
[30] E. Laine, L. Martinez, A. Blondel, and T. E. Malliavin, Biophys. J. **99**, 2264 (2010).
[31] M. I. Stefan, D. P. Marshall, and N. Le Novere, PLoS One **7**, e29406 (2012).
[32] A. Malmendal, J. Evenas, S. Forsen, and M. Akke, J. Mol. Biol. **293**, 883 (1999).
[33] M. A. Shea and B. R. Sorensen, Biochemistry **37**, 4244 (1998).
[34] B. R. Sorensen, L. A. Faga, R. Hultman, and M. A. Shea, Biochemistry **41**, 15 (2002).
[35] O. R. Jaren, J. K. Kranz, B. R. Sorensen, A. J. Wand, and M. A. Shea, Biochemistry **41**, 14158 (2002).
[36] Y. Ye, H. W. Lee, W. Yang, S. Shealy, and J. J. Yang, J. Am. Chem. Soc. **127**, 3743 (2005).
[37] M. Lepsik and M. J. Field, J. Phys. Chem. B **111**, 10012 (2007).
[38] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Chem. Phys. Lett. **156**, 472 (1989).
[39] M. Sprik and G. Ciccotti, J. Chem. Phys. **109**, 7737 (1998).
[40] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden, Chem. Phys. Chem. **6**, 1809 (2005).
[41] I. Coluzza, M. Sprik, and G. Ciccotti, Mol. Phys. **101**, 2885 (2003).
[42] J. Schlitter and M. Klahn, J. Chem. Phys. **118**, 2057 (2003).
[43] W. K. den Otter, J. Chem. Phys. **112**, 7283 (2000).
[44] J. Schlitter, Eur. Phys.: J. Spec. Top. **200**, 91 (2011).
[45] J. Schlitter and M. Klahn, Mol. Phys. **101**, 3439 (2003).
[46] E. Darve and A. Pohorille, J. Chem. Phys. **115**, 9169 (2001).

[47] W. K. den Otter and W. J. Briels, J. Chem. Phys. **109**, 4139 (1998).
[48] V. Ovchinnikov, M. Karplus, and E. Vanden-Eijnden, J. Chem. Phys. **134**, 085103 (2011).
[49] S. Shirran, P. Garnaud, S. Daff, D. McMillan, and P. Barran, J. R. Soc., Interface **2**, 465 (2005).
[50] A. Lewit-Bentley and S. Rety, Curr. Opin. Struct. Biol. **10**, 637 (2000).
[51] P. W. Rose, B. Beran, C. X. Bi, W. F. Bluhm, D. Dimitropoulos, D. S. Goodsell, A. Prlić, M. Quesada, G. B. Quinn, J. D. Westbrook, J. Young, B. Yukich, C. Zardecki, H. M. Berman, and P. E. Bourne, Nucleic Acids Res. **39**, D392 (2011).
[52] P. Ren and J. W. Ponder, J. Phys. Chem. B **107**, 5933 (2003).
[53] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[54] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[55] V. C. Still, A. Tempezvk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
[56] C. Chen and Y. Xiao, Bioinformatics **24**, 659 (2008).
[57] H. X. Lei and Y. Duan, J. Mol. Biol. **370**, 196 (2007).
[58] H. X. Lei, C. Wu, H. G. Liu, and Y. Duan, Proc. Natl. Acad. Sci. USA **104**, 4925 (2007).
[59] D. Jiao, C. King, A. Grossfield, T. A. Darden, and P. Y. Ren, J. Phys. Chem. B **110**, 18553 (2006).
[60] R. Schmid, A. M. Miah, and V. N. Sapunov, PhysChemChemPhys **2**, 97 (2000).
[61] D. Buntinas, C. Coti, T. Herault, P. Lemarinier, L. Pilard, A. Rezmerita, E. Rodriguez, and F. Cappello, Future Gener. Comput. Syst. **24**, 73 (2008).
[62] C. Chen and Y. Xiao, J. Comput. Chem. **31**, 1368 (2010).
[63] T. P. Straatsma and J. A. McCammon, J. Chem. Phys. **95**, 1175 (1991).
[64] A. Cornish-Bowden, J. Biosci. **27**, 121 (2002).
[65] Y. Komeiji, Y. Ueno, and M. Uebayasi, FEBS Lett. **521**, 133 (2002).
[66] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).