Protein−Ligand Binding Free Energy Calculation by the Smooth Reaction Path Generation (SRPG) Method

Yoshifumi Fukunishi,*[,†,‡] Daisuke Mitomo,[§] and Haruki Nakamura[†,||]

Biomedicinal Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan, Pharmaceutical Innovation Value Chain, BioGrid Center Kansai, 1-4-2 Shinsenri-Higashimachi, Toyonaka, Osaka 560-0082, Japan, Japan Biological Informatics Consortium (JBIC), 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan, and Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Received June 16, 2009

We developed a new molecular dynamics simulation method for protein−ligand binding free energy calculation in an explicit water model. This method consists of three steps: (1) generation of a compound dissociation path starting from a stable protein−compound complex structure, (2) calculation of the free energy surface along the dissociation path, and (3) calculation of the free energy surface of a small area around the protein−compound complex structure, which is a free energy minimum. The protein−compound binding free energy is estimated from the information obtained by the above three steps. This method was applied to a small system, a 18-crown-6 ether with its ligand ion, and a realistic system consisting of a target protein with its inhibitor. This approximation worked well; the protein−inhibitor dissociation was successfully performed, and the binding free energies were calculated.

1. INTRODUCTION

Estimation of protein−compound binding free energy (ΔG) is important for structure-based drug development, and many protein−compound docking programs and scoring functions have been reported.[1−7] The protein−compound docking programs and scoring functions are very fast computationally and, thus, are useful in in-silico (virtual) drug screening. However, the docking scores are not sufficiently precise to achieve the desired binding free energy error of 2−3 kcal/mol.[8−11] The molecular-mechanics Poisson−Boltzmann surface-area (MMPBSA) method[12,13] and the recent linear interaction energy (LIE) method[14−18] have succeeded in reproducing the trend of ΔGs for a single target protein. In the MMPBSA method, hydrophobic interaction, which is an entropy term, is evaluated by using the surface area, which provides only a rough, imprecise approximation. In the LIE method, the ΔG is evaluated based on the enthalpy term. This term consists of the average van der Waals energy and average electrostatic energy. In the framework of statistical physics, the binding free energy is calculated based on the partition function of the system obtained from the canonical ensemble, which consists of numerous structures. The docking score, in contrast, is calculated based on a single structure of the system. Theoretically, we cannot calculate the precise binding free energy by using the scoring function.

There have been several reports on protein−compound docking and free energy calculation by the molecular dynamics (MD) simulation. Even if the protein−ligand complex structure is unknown, ab initio MD docking simulations show the protein−ligand complex structures and the free energy landscapes.[19,20] The generalized ensemble methods have been adopted for wide conformational searches.[21−24] The multicanonical molecular dynamics (McMD) simulation has been applied to protein−ligand docking problems.[19,20] The McMD method calculates the density of state of the system by a long-time MD simulation. This approach is theoretically exact, but generally speaking, it is very time-consuming.

An approach faster than the McMD method would thus be desirable. In an explicit water model, if the protein−ligand complex structure is known, the binding free energy and the potential of mean force (PMF) along the dissociation path can be obtained by using the filling potential (FP) method,[24] meta dynamics method,[25,26] and Jarzynski's method.[27−29] We previously proposed the FP method,[24] which generates a reaction path (dissociation path) of the ligand and calculates the free energy surface along the path based on ab initio MD simulation. The FP method is an umbrella potential sampling method that enables the ligand molecule to drift from its local minima automatically. Without setting the reaction coordinates a priori, this method searches for and determines suitable reaction coordinates by the successive generation of umbrella potentials from trajectory analysis. The weighted histogram analysis of these trajectories gives the binding free energy. This method has been applied to a complex of thermolysin and its inhibitors. The other trend is the application of Jarzynski's equation. In this method, just as in the FP method, a harmonic potential that restrains the ligand at a particular position moves slowly and leads the ligand from the binding state to the dissociation state, and the free energy profile is calculated.

* To whom correspondence should be addressed. Tel: +81-3-3599-8290. Fax: +81-3-3599-8099. E-mail: y-fukunishi@aist.go.jp.
† National Institute of Advanced Industrial Science and Technology (AIST).
‡ BioGrid Center Kansai.
§ Japan Biological Informatics Consortium (JBIC).
|| Osaka University.

10.1021/ci9002156 CCC: $40.75
Published on Web 07/29/2009

In the current study, we proposed a new approach by introducing modified FP method for ΔG calculation.

In the usual FP method, the convergence of the weighted histogram analysis method (WHAM) is time-consuming, and it is difficult to obtain trajectories with enough overlap to converge the WHAM. We therefore proposed a modified FP in which the thermodynamic integration (TI)[30] method is used in place of the WHAM.

2. MATERIALS AND METHODS

The current method calculates the free energy difference between a protein−compound complex structure (binding state) and its isolated structure (unbound state), and here, the binding state is a free energy minimum. This method consists of three steps: (1) generation of a compound dissociation path starting from a stable protein−compound complex structure, (2) calculation of the free energy surface along the dissociation path, and (3) calculation of the free energy surface around the protein−compound complex structure, which is a free energy minimum.

The ΔG is

$$\Delta G = -k_B T \ln \frac{P_B}{P_U} \quad (1)$$

where k_B, T, P_B, and P_U are the Boltzmann constant, temperature, and un-normalized probabilities of the bound and unbound states, respectively. The un-normalized probabilities are given by

$$P_B = \int_{R_B} \exp(-\beta G(\mathbf{r}))d\mathbf{r} \quad (2)$$

$$P_U = \int_{R_U} \exp(-\beta G(\mathbf{r}))d\mathbf{r} \quad (3)$$

where G($\mathbf{r}$), β, R_B, and R_U are the free energy at position $\mathbf{r}$, $1/k_BT$, the region of the bound state, and the region of the unbound state, respectively.

If the major contribution of P_B is given by the integral of eq 2 in a small area around the free energy minimum ($\mathbf{r}_0$), the $G(\mathbf{r})$ could be approximated by a harmonic potential as follows:

$$G(\vec{r}) = G(\vec{r}_0) + \left(\frac{k_x}{2}\Delta x^2 + \frac{k_y}{2}\Delta y^2 + \frac{k_z}{2}\Delta z^2\right) \quad (4)$$

where

$$\vec{r} = \vec{r}_0 + \Delta\vec{r} \quad (5)$$

$$\Delta\vec{r} = (\Delta x, \Delta y, \Delta z) \quad (6)$$

In eqs 4 and 5, the xyz axes represent the principal axes of the free energy surface. Thus,

$$P_B = \int_{V_B} \exp\left(-\beta\left(G(\mathbf{r}_0) + \frac{k_x}{2}x^2 + \frac{k_y}{2}y^2 + \frac{k_z}{2}z^2\right)\right)dxdydz \quad (7)$$

where V_B represents the binding region. The P_u value is calculated as

$$P_U = \int_0^R 4\pi\mathbf{r}^2 \exp(-\beta G(\mathbf{r}_\infty))d\mathbf{r} = \frac{4\pi}{3}R^3 \times \exp(-\beta G(\mathbf{r}_\infty)) = V_0 \exp(-\beta G(\mathbf{r}_\infty)) \quad (8)$$

where V_0 represents the volume for one compound at 1 M concentration.

By integrating eqs 7 and 8 into eq 1 and setting the radius of the bound region to ∞ for the purpose of simplicity, we get

$$\Delta G = G(\mathbf{r}_0) - G(\mathbf{r}_\infty) - k_B T \ln \int_{-\infty}^{\infty} \exp\left(-\beta\frac{k_x}{2}x^2\right)dx \times \int_{-\infty}^{\infty} \exp\left(-\beta\frac{k_y}{2}y^2\right)dy \int_{-\infty}^{\infty} \exp\left(-\beta\frac{k_z}{2}z^2\right)dz/V_0 \quad (9)$$

Here k_x, k_y, and k_z are the force constants obtained from the free energy surface. At 1 M concentration, the volume for one compound, V_0, is 1661 Å^3. The analytical form of eq 9 is

$$\Delta G = G(\mathbf{r}_0) - G(\mathbf{r}_\infty) - k_B T \ln \frac{\frac{\sqrt{\pi}}{\sqrt{\beta k_x/2}}\frac{\sqrt{\pi}}{\sqrt{\beta k_y/2}}\frac{\sqrt{\pi}}{\sqrt{\beta k_z/2}}}{1661} \quad (10)$$

If the free energy surface around $\mathbf{r}_0$ is isotropic, $k_x = k_y = k_z = k$ in eq 10.

$G(\mathbf{r}_0)$ could be obtained from the PMF along the dissociation path, and k could be obtained from the free energy surface around the stable protein−compound structure ($\mathbf{r}_0$). Finally, we must know the $G(\mathbf{r}_0) - G(\mathbf{r}_\infty)$ to evaluate ΔG. We call this method the smooth-reaction path generation (SRPG) method, and describe it in detail below.

2.1. Generation of the Dissociation Path. The dissociation path is generated by a rough MD simulation. The ligand dissociation path is generated by using the FP method at high temperature in a vacuum. The protein−compound complex structure is simulated with a short cutoff of 1−5 interactions (van der Waals and Coulomb interactions) to reduce the CPU time. The atomic charges of the compound are set to zero to reduce the protein−compound attractive interaction energy. The trajectory of the ligand should not be smooth. The trajectory of the compound is the initial guess at a dissociation path. One of the atoms of the ligand molecule is selected as a landmark atom to represent the trajectory, and the coordinates of the n-th position of the landmark atom are denoted as $\mathbf{p}^0(n)$. Let M be the number of trajectory frames. The initial guess at a dissociation path is $\{\mathbf{p}^0(n); n = (1, 2, 3,..., M)\}$.

2.2. Smoothing of the Dissociation Path. A Legendre fitting method generates a smooth reaction path based on the initial guess made in the previous step. To perform the TI method in the next step, we need a smooth reaction path instead of $\{\mathbf{p}^0(n); n = (1, 2, 3,..., M)\}$. The initial coordinates and the final coordinates are selected from the trajectory of the compound. The one-parameter reaction path $\mathbf{p}(t)=\{p_x(t), p_y(t), p_z(t); t = [0, 1]\}$ is

$$\begin{cases} p_x(t) = \sum_{i=0}^{L} c_x^i P_i(t) \\ p_y(t) = \sum_{i=0}^{L} c_y^i P_i(t) \\ p_z(t) = \sum_{i=0}^{L} c_z^i P_i(t) \end{cases} \quad (11)$$

where $P_i(t)$ and t are the i-th Legendre function and a parameter ($0 \le t \le 1$). The L value controls the curvature of the reaction path. If $L = 1$, the path is linear. The Legendre function is complete so that a linear combination of the Legendre function can represent any kind of function.

The points $\mathbf{p}(0)$ and $\mathbf{p}(1)$ are equal to the initial and the final coordinates for any values of the coefficients, respectively. The optimal coefficients are determined by a Monte Carlo simulation. For practical calculation, a set of discrete points on the path is necessary. N points are generated along $\mathbf{p}(t)$, and the n-th point is $\mathbf{p}(n/N)$. Let D be distance between two points. The S value

$$S = \sum_{m}^{M} \sum_{n}^{N} D(\vec{p}(n'N), \vec{p}^{\,0}(m))^2 \quad (12)$$

is calculated for various values of the coefficients c_x^i, c_y^i, and c_z^i. The optimal parameter set $\{c_x^i, c_y^i, c_z^i, i = 1, 2,.., L\}$, which minimizes the S value, is determined by a Monte Carlo calculation.

2.3. Free Energy Calculation Based on the Smooth Dissociation Path. The landmark atom is restrained at each point on the dissociation path by an umbrella potential, which is a harmonic potential. The force acting on the landmark atom is calculated as

$$\langle \vec{F}(\vec{r}) \rangle_{V_0} = \langle \vec{F} \rangle_{V-\Delta V, \Omega(\mathbf{r})} = \frac{\int_{\Omega(\mathbf{r})} \vec{F} \mathrm{e}^{-\beta(V-\Delta V)} \mathrm{d}\mathbf{r}}{\int_{\Omega(\mathbf{r})} \mathrm{e}^{-\beta(V-\Delta V)} \mathrm{d}\mathbf{r}} = \frac{\langle \vec{F} \mathrm{e}^{\beta \Delta V} \rangle_{V,\Omega(\mathbf{r})}}{\langle \mathrm{e}^{\beta \Delta V} \rangle_{V,\Omega(\mathbf{r})}} \quad (13)$$

where $\mathbf{F}$, $\mathbf{r}$, V_0, V, ΔV, $\Omega(\mathbf{r})$, $\langle\ \rangle_{V,\Omega(\mathbf{r})}$ represent the force, the position of the landmark atom of ligand, the potential without the umbrella potential, the potential with the umbrella potential ($V = V_0 + \Delta V$), the umbrella potential, the small region around the position $\mathbf{r}$, and the average in the region $\Omega(\mathbf{r})$ with potential V, respectively.

The free energy surface (PMF) is calculated using the TI method as follows:

$$G(R) = \int_0^R \langle \vec{F}(\vec{r}) \rangle \cdot \mathrm{d}\vec{r} \quad (14)$$

Thus, $G(\mathbf{r}_0)$ and $G(\mathbf{r}_\infty)$ are given by the PMF along the reaction path, and the force constant k could be given by a free energy calculation around the free energy minimum.

To calculate the k value, $G(\mathbf{r})$ around $\mathbf{r}_0$ is calculated. Namely, $G(\mathbf{r})$ values for $\mathbf{r} = \mathbf{r}_0 + n_x \Delta \mathbf{r}_x$, $\mathbf{r} = \mathbf{r}_0 + n_y \Delta \mathbf{r}_y$, and $\mathbf{r} = \mathbf{r}_0 + n_z \Delta \mathbf{r}_z$, where n_x, n_y, $n_z = [-L_\mathrm{m}, L_\mathrm{m}]$, and $\Delta \mathbf{r}_x$, $\Delta \mathbf{r}_y$, and $\Delta \mathbf{r}_z$ are small vectors along the x, y, and z axis, respectively. In the current study, L_m, and the length of vectors ($\Delta \mathbf{r}_x$, $\Delta \mathbf{r}_y$, $\Delta \mathbf{r}_z$) were set to 10 and 0.1 Å, respectively. The $G(\mathbf{r})$ and k values are fitted to eq 4 by least-squares fitting.

2.4. Filling Potential (FP) Model. The FP method was previously reported by our group.[24] Here, we provide a brief explanation of this method. The FP method is an umbrella potential sampling method that enables the ligand molecule to drift from its local minima automatically. The umbrella potential is a combination of Gaussian-type repulsive potentials, which are located on the trajectory of the ligand. Without setting the reaction coordinates a priori, this method searches for and determines suitable reaction coordinates by the successive generation of umbrella potentials based on its trajectory analysis. The method for obtaining the reaction path is based on the concept of the Taboo search.

To mimic a self-avoiding trajectory, the FP method sets a repulsive potential on the previous position of the ligand in the MD simulation. If the repulsive potential is not high enough to avoid a visit by the ligand, the ligand can visit the same position again. An additional repulsive potential is then added to the previous repulsive potential, until the ligand can no longer visit the same position. Since the FP method constructs the potential based on a set of the previous positions of the ligand, for longer simulations, a larger allocation of memory is required. To reduce the memory requirement, the method uses the coordinates of one atom of the ligand molecule to construct the potential. The selected atom is designated as "landmark".

The umbrella potential for FP (V_FP) consists of the repulsive potentials (V_rep) and a position-restraint potential (V_res). The umbrella potential for the n-th simulation (V_FP^n) is

$$V_\mathrm{FP}^n = \sum_{k=1}^{n-1} V_\mathrm{rep}^k + V_\mathrm{res}^n \quad (15)$$

where V_rep^k is a repulsive potential constructed from all previous simulation, and V_res^n is a position restraint potential that restrains the ligand around the initial coordinates of the n-th simulation. Both V_rep^k and V_res^n act on the landmark atom only.

Let $X(t)_\mathrm{landmark} = \{x(t), y(t), z(t)\}_\mathrm{landmark}$ be a set of x, y, z coordinates at time t of the landmark atom. Let $X^k = (x^k, y^k, z^k)$ be the final coordinates of the landmark atom of the k-th simulation. V_rep^k is a Gaussian-type function defined as follows:

$$V_\mathrm{rep}^k(X(t)_\mathrm{landmark}) = c_1(k) \exp[c_2(X(t) - X^k) \cdot (X(t) - X^k)] \quad (16)$$

where $c_1(k)$ and c_2 determine the height and the width of the Gaussian potential.

V_res^n works to localize the ligand around a specific position. Suitable choices of V_res^n can retain the overlap of trajectories between the n-th and the $(n + 1)$-th simulations. Let $D^k(t)$ be a distance of the landmark atom between that at time t and the final position of the k-th simulation

$$D^k(t) = \sqrt{(x(t) - x^k)^2 + (y(t) - y^k)^2 + (z(t) - z^k)^2} \quad (17)$$

V_res^n is a kind of CAP-type potential defined as follows:

$$V^n_{\mathrm{res}}(X(t)_{\mathrm{landmark}}) = \begin{cases} \frac{1}{2}k^n D^{n-1}(t) & \text{when } D^{n-1}(t) \geq D_0 \\ 0 & \text{when } D^{n-1}(t) < D_0 \end{cases} \quad (18)$$

where k^n is a force constant and D_0 is a radius of this half harmonic potential. For $n = 1$, x^k, y^k, and z^k are set to the initial coordinates of the landmark atom in the first simulation.

The first simulation is performed on the umbrella potential $V^1_{\mathrm{FP}} = V^1_{\mathrm{res}}$. The ligand molecule moves around the initial coordinates. The second simulation is performed on $V^2_{\mathrm{FP}} = V^1_{\mathrm{rep}} + V^2_{\mathrm{res}}$. The ligand again moves around the final coordinates of the first simulation. The third simulation is performed on $V^3_{\mathrm{FP}} = V^1_{\mathrm{rep}} + V^2_{\mathrm{rep}} + V^3_{\mathrm{res}}$. The ligand moves around the final coordinates of the second simulation and feels a repulsive potential from the first coordinates of the first and the second simulations. This iterative procedure is performed until the ligand reaches the dissociated state. The parameter c_2 of eq 16 corresponds to the width of the umbrella potential. The width of the umbrella potential must be smaller than the system size. For example, a constant umbrella potential, which covers the whole system, is useless. The height of the umbrella potential must be small to retain the overlap of trajectories of the different states.

2.5. Computational Model. We built two systems; a 18-crown-6 ether with its ligand, a K^+ ion,[31] and a structure consisting of streptavidin complexed with biotin (PDB id 1stp; see Figures 1 and 2).

The 3D coordinates of the 18-crown-6 ether were generated by the Chem3D program (Cambridge Software, Cambridge, MA). We used the general AMBER force field (GAFF),[32] and the molecular topology files were generated by tplgeneL/myPresto. The atomic charges of these molecules were determined by the restricted electrostatic point charge (RESP) procedure using HF/6-31G*-level quantum chemical calculations with the program Gaussian 98.[33] The crown ether + K^+ structure was placed in a unit cell of lattice. The TIP3P water model was adopted for water molecules.[34] The system consisted of 1495 atoms (42 crown ether, 1 K^+, 2 Cl^-, 1 Na^+, and 1449 water atoms). The cell size was optimized by NPT MD simulation at 1 atm pressure with the Andersen method.[35] The simulation time was 600 ps with a time step of 1.5 fs, the 1−5 interaction was calculated by the PME method with a cutoff distance of 8 Å, and the average cell size was 24.66 Å × 24.66 Å × 24.66 Å.

For the protein−inhibitor system, all of the Asp, Glu, Arg, and Lys residues were treated as being charged. The protein−inhibitor structure was placed in a sphere (radius = 25 Å) consisting of water molecules. We call this sphere the water sphere. The final system with biotin thus consisted of 6643 atoms (1744 protein, 31 inhibitor, 4 Cl^-, 7 Na^+, and 4857 water atoms). We used this structure as the starting conformation of the SRPG simulation.

Figure 2 shows the chemical structure of biotin. The atomic charges of the ligands were determined by the restricted electrostatic point charge (RESP) procedure using HF/6-31G*-level quantum chemical calculations with the program Gaussian 98.[33] The force field parameters for the protein were the AMBER parm99 force field.[36] That for biotin was the general AMBER force field (GAFF).[32] The TIP3P water model was adopted for water molecules.[34]

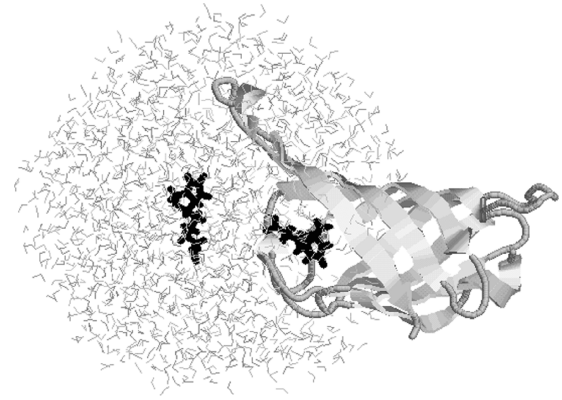

Figure 1. Protein−inhibitor system calculated in this study (streptavidin with its inhibitor biotin).

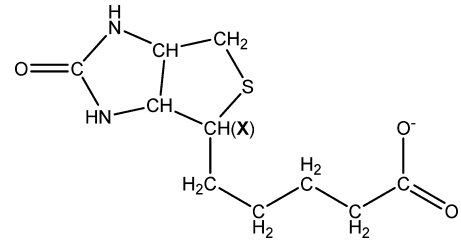

Figure 2. Inhibitor of the protein, biotin. The carbon atom (X) is the landmark atom.

3. RESULTS

The computer program myPresto, version 4 (myPresto, http://medals.jp/myPresto/index.html (accessed July 18, 2009) and myPrestoManagement, http://presto.protein.osaka-u.ac.jp/myPresto4/index_e.html (accessed July 18, 2009)) was used for the simulation.[24]

3.1. Free Energy Landscape and the ΔG of Crown Ether with K^+. We applied the SRPG method to the 18-crown-6 ether + K^+ ion in water. The ring structure of the 18-crown-6 is on the $y-z$ plane. A simple reaction path was adopted in this case; the linear dissociation path was set to the crown (x axis). The SRPG procedure was as follows:

Step 1. For the $\mathbf{p}(n/N)$, where $n = 1,..., 72$, the K^+ atom was placed on the $\mathbf{p}(n/N)$. Each system was optimized by energy minimization with position restraint potential onto the heavy atoms of the crown ether and the K^+ atom.

Step 2. The MD simulation of each system was performed at 300 K for 300 ps for data sampling, after an MD simulation of 300 ps for equilibration. The average force acting on the landmark atom was calculated for each $\mathbf{p}(n/N)$.

Step 3. The PMF was constructed from the average force calculated in step 2.

Step 4. The free energy surface around the stable binding state was calculated. Δr_x, Δr_y, and Δr_z were set to 0.1 Å, and the free energy surface was calculated in the same way as in step 2.

Step 5. The PMF and ΔG were calculated based on eqs 9 and 10.

The SHAKE method was used to constrain covalent bonds between heavy and hydrogen atoms in any molecule in the system.[37] The electrostatic interactions were computed by a particle-mesh Ewald method without truncation.[38] The unit time step for the MD simulation was set to 1.5 fs.

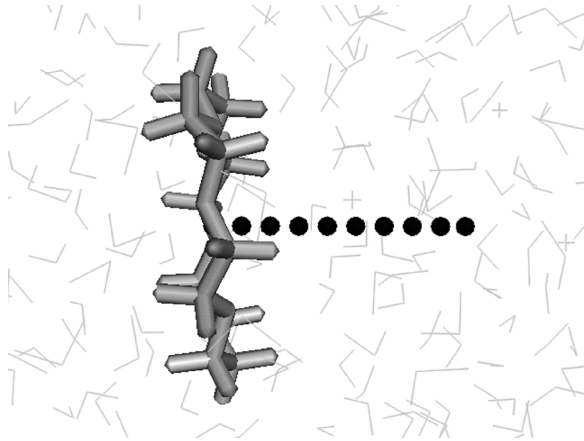

Figure 3. $x-z$ projection of the reaction path of the 18-crown-6 with K^+ system.

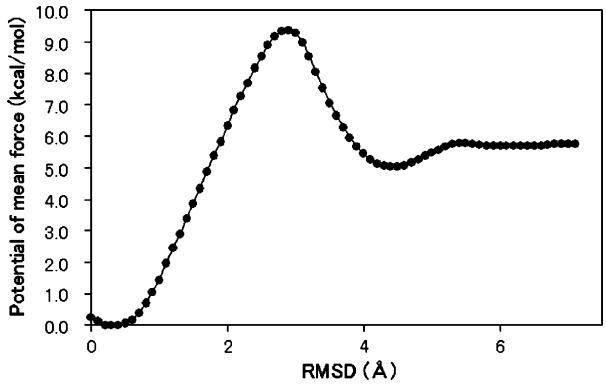

Figure 4. Potential of mean force along the reaction path of the 18-crown-6 with K^+ system.

Figure 3 shows the trajectories of dissociation and the smooth dissociation path. Figure 4 shows the PMF of this system. There are two major minima at 0.4 and at 4.4 Å, and these minima are separated by a free energy barrier at 2.9 Å. The profile of the PMF is very close to that of the PMF reported in a previous study. In the current study, the free energy at 0.4 Å is 0 kcal/mol, the height of the free energy barrier at 2.9 Å is 9.38 kcal/mol, and the second free energy minima at 4.4 Å is 5.03 kcal/mol. The free energy profile was similar to those of the previous studies.[31,39,40] These results show that the PMF calculation worked well and is reliable.

Figure 5 shows the free energy surface around the bound state (rmsd = 0 Å). The free energy surfaces along the x, y, and z axes were calculated with the same condition used to depict the PMF. The ring structure of the 18-crown-6 is on the $y-z$ plane and the dissociation path is perpendicular to the crown (x axis). Thus, the principal component axes around the bound state should be the x, y, and z axes. The results show that the free energy surface around the bound state is anisotropic, and the k_x, k_y, and k_z values of eq 9 were 2.06, 10.08, and 10.01 kcal/(mol Å^2), respectively. The values of $G(\mathbf{r}_0)$ and $G(\mathbf{r}_\infty)$ were 0 and 5.75 kcal/mol, respectively. The calculated binding free energy was −0.71 kcal/mol, which is somewhat different from the experimental value of −2.8 to −3.0 kcal/mol.[41] In addition, the P_B of eq 6 must be doubled because the binding sites are located on both sides of the ring.

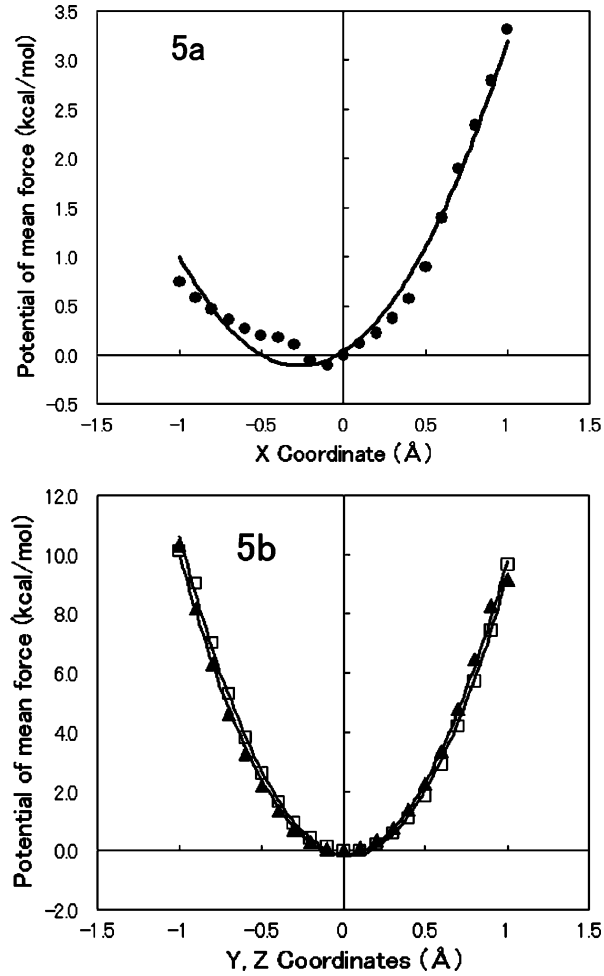

Figure 5. Free energy surface around the bound state of the 18-crown-6 with K^+ system. Circles, open squares, and triangles represent the free energies along the x, y, and z axes.

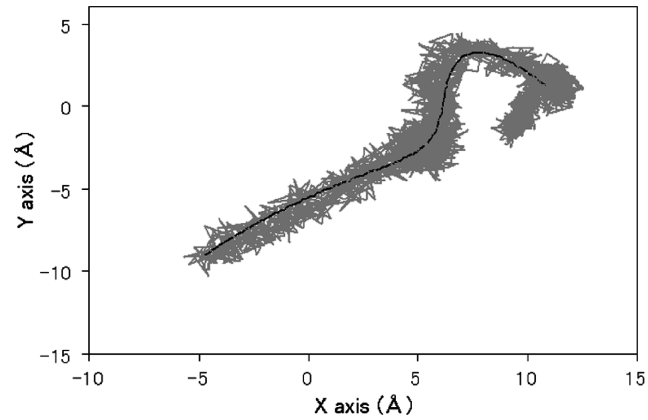

Figure 6. $x-y$ projection of the smooth reaction path generated for the streptavidin with biotin system. The gray line represents the trajectory obtained by the FP method in a vacuum and the black curve represents the smooth reaction path generated by the SRPG method.

This simple example showed that the SRPG method actually works and the results are compatible to those in previous studies.

3.2. Free Energy Landscape and the ΔG of a Protein with Its Inhibitor. We applied the SRPG method to a protein and its inhibitor in water. The SRPG procedure was almost the same as that of the crown ether system.

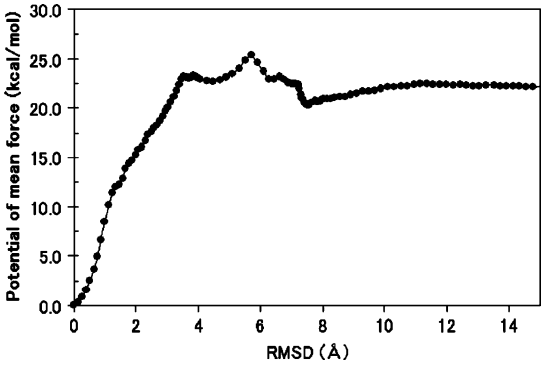

Figure 7. Potential of mean force of the streptavidin with biotin system along the smooth reaction path. rmsd = 0 Å represents the bound state, and rmsd = 15 Å represents the dissociated state.

Step 1. The whole system was energy minimized with a position restraint potential on all heavy atoms of the protein. A preparatory MD simulation of 500 psec at 300K was performed to equilibrate the system.

Step 2. The trajectory of the landmark atom was sampled by the 200 ps MD simulation at 300 K with the umbrella potential V_{FP}. In this case, the X atom of the ligand was selected as the landmark atom (see Figure 2). For the first MD, $V_{rep} = 0$.

Step 3. The parameters of V_{FP} were determined by the last coordinate of the landmark atom.

Step 4. V_{FP} was reconstructed. Then, we returned to step 1. This procedure was repeated until the ligand molecule reached the dissociated state.

Step 5. After steps 1−4 were completed, the smooth dissociation path was constructed from the trajectories of the X atom of the ligand. The L value was set to 4, and the N value was set to 112.

Step 6. For the $\mathbf{p}(n/N)$ where $n = 1,..., 112$, the landmark atom was placed on the $\mathbf{p}(n/N)$. Each system was optimized by energy minimization with position restraint potential onto the heavy atoms of the protein and the landmark atom.

Step 7. The MD simulation of each system was performed at 300 K of 750 ps for data sampling, after an MD simulation of 1750 ps for equilibration. The average force acting on the landmark atom was calculated for each $\mathbf{p}(n/N)$.

Step 8. The PMF was constructed from the average force calculated in step 7.

Step 9. The free energy surface around the stable binding state was calculated. Δr_x, Δr_y, and Δr_z were set to 0.1 Å, and the free energy surface was calculated as in steps 7−8.

Step 10. The PMF and ΔG were calculated based on eq 10.

The SHAKE method was used to constrain covalent bonds between heavy and hydrogen atoms in any molecule in the system.[37] The electrostatic interactions were computed by a fast-multipole method without truncation.[42] The unit time step for the MD simulation was set to 1.5 fs. The radius for sampling in eq 10 was set to 0.3 Å.

Figure 7 is the PMF of the free energy landscape on the $\text{rmsd}^{\text{inhibitor}}$ axis, where the $\text{rmsd}^{\text{inhibitor}}$ is the rmsd of heavy atoms of the inhibitor from the native complex structure. Figure 7 shows that a free-energy barrier is clearly seen at an $\text{rmsd}^{\text{inhibitor}}$ of 6 Å. There are two basins: the natively bound ligand basin and unbound ligand basin. The bottom of the natively bound ligand basin was at an $\text{rmsd}^{\text{inhibitor}}$ of 0 Å, and the bottom of the unbound ligand basin was at an $\text{rmsd}^{\text{inhibitor}}$ of 7.5 Å. The barrier height was 5.1 kcal/mol measured from the bottom of the unbound ligand basin and was 25.4 kcal/mol from the bottom of the natively bound ligand basin.

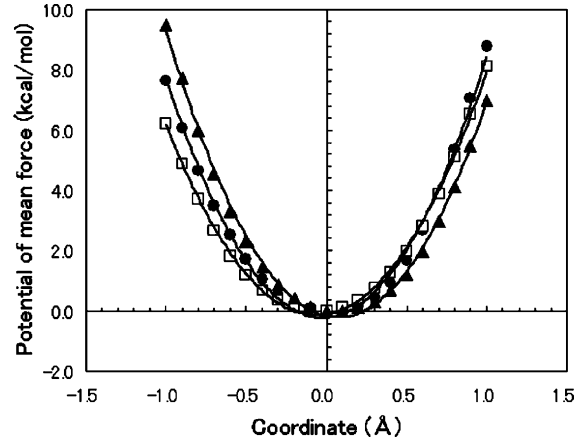

Figure 8. Free energy surface around the bound state of the streptavidin with biotin system. Circles, open squares, and triangles represent the free energies along the x, y, and z axes.

Figure 8 shows the free energy surface around the bound state (rmsd = 0 Å). The free energy surfaces along the x, y, and z axes were calculated under the same conditions used to depict the PMF. The results show that the free energy surface around the bound state is isotropic, and the average k value of eq 4 was 7.90 kcal/(mol Å^2). The k_x, k_y, and k_z values were 8.28, 7.15, and 8.27 (mol Å^2), respectively. The values of $G(\mathbf{r}_0)$ and $G(\mathbf{r}_\infty)$ were 0 and 22.1 kcal/mol, respectively.

The calculated binding free energy obtained from eq 10 was −16.5 kcal/mol, which is close to the experimental value of −18.3 kcal/mol.[4] The free energy surface around the binding state was slightly anisotropic. If $k = k_x$, $k = k_y$, or $k = k_z$, the ΔG values were −16.5, −16.6, −16.5 kcal/mol, respectively. Thus, the error in the estimation of the ΔG value due to anisotropy should be small. This realistic example showed that the SRPG method actually works and the accuracy was comparable to that observed experimentally.

4. DISCUSSION

To obtain the PMF, the force vectors were obtained from 50−100 individually calculated trajectories. In addition, the number of water molecules could be different for each system, since we only require the force acting on the landmark atom of the ligand molecule. Thus the SRPG method is suitable for parallel simulation as same as the MP-CAFEE method that utilizes the histogram method and a perturbation method.[43] The dissociation path must be generated from a single MD simulation, but the simulation is not time-consuming. The most time-consuming step is the calculation of the force vectors, and this step was parallelized in the current study. The water molecules, which are far from the ligand molecule, should not effect the force acting on the landmark atom. Thus such water molecules could be ignored in the calculation and the system size could be reduced to speed up the simulation time.

The surface plasmon resonance (SPR) experiment showed that the processes by which ligands become dissociated from

and adsorbed to a target protein take several seconds in general.[44] These processes are as slow as the protein folding.[45] Thus, we could expect large-scale motions of proteins (domain motion, side-chain rotation, local and global unfolding and folding) during the process of the dissociation of the ligand molecule from the target protein. Compared to the actual dissociation process, the dissociation simulation in the current study was so fast that the computational error could be serious. Furthermore, the position restraint of atoms of the target protein reduced the flexibility of the target protein. One of the serious problems is the motions of water molecules. In the case of a deep ligand binding pocket, it is difficult for the water molecules to come in the pocket instead of the ligand during the ligand dissociation process.[24] In the current study, the ligand binding pocket was not very deep. To apply the SRPG method to a deep ligand binding pocket, some amino-acid residues around the pocket should be removed to increase the mobility of the water molecules.

5. CONCLUSIONS

We proposed the smooth reaction path generation (SRPG) method, which is a new MD simulation procedure for binding free energy calculation. First, starting from a stable protein−compound complex structure, a rough compound dissociation path was generated by means of the filling potential method, which was previously developed by our group. Second, a smooth reaction path was constructed based on the rough dissociation path. Third, the free energy surface along the smooth reaction path was calculated by using a type of thermodynamic integration method. Fourth, the free energy surface around the stable protein−compound complex structure was calculated. And last, the binding free energy was calculated from the free energy surface along the dissociation path and the free energy surface around the complex structure.

The SRPG method was applied to two systems: a 18-crown-6 ether with its ligand ion and a target protein with its inhibitor. The binding free energies were successfully reproduced based on the ab initio MD simulations with an explicit water model. The SRP method was suitable for parallel calculation. Also, the potential of mean force values were close to those obtained in previous works. This procedure should be useful in drug designs that require the examination of many different kinds of ligands.

ACKNOWLEDGMENT

This work was supported by grants from the New Energy and Industrial Technology Development Organization of Japan (NEDO) and the Ministry of Economy, Trade, and Industry (METI) of Japan.

REFERENCES AND NOTES

(1) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule−ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
(2) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
(3) Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
(4) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins* **1998**, *33*, 367–382.
(5) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195–202.
(6) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM: A new method for structure modeling and design: Application to docking and structure prediction from the disordered native conformation. *J. Comput. Chem.* **1994**, *15*, 488–506.
(7) Fukunishi, Y.; Mikami, Y.; Nakamura, H. Similarities among receptor pockets and among compounds: Analysis and application to in silico ligand screening. *J. Mol. Graph. Modell.* **2005**, *24*, 34–45.
(8) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring function. *J. Med. Chem.* **2006**, *49*, 5912–5931.
(9) Fukunishi, Y.; Mikami, Y.; Kubota, S.; Nakamura, H. Multiple target screening method for robust and accurate in silico ligand screening. *J. Mol. Graph. Modell.* **2005**, *25*, 61–70.
(10) Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. A knowledge-based energy function for protein−ligand, protein−protein, and protein−DNA complexes. *J. Med. Chem.* **2005**, *48*, 2325–2335.
(11) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.
(12) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Srivasan, J.; Case, D. A.; Cheatam, T. E., III. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.
(13) Obiol-Pardo, C.; Rubio-Martinez, J. Comparative evaluation of MMPBSA and XSCORE to compute binding free energy in XIAP-peptide complexes. *J. Chem. Inf. Model.* **2007**, *47*, 134–42.
(14) Almlof, M.; Carlsson, J.; Åqvist, J. Improving the accuracy of the linear interaction energy method for salvation free energies. *J. Chem. Theory. Comput.* **2007**, *3*, 2162–2175.
(15) Tiunge, B. A.; Rajamani, R.; Baxter, E. W.; Reitz, A. B.; Reynolds, C. H. Linear interaction energy models for β-serectase (BACE) inhibitors: role of van der Waals, electrostatic, and continuum-solvation terms. *J. Mol. Graph. Modell.* **2006**, *24*, 475–484.
(16) Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput-Aided. Mol. Des.* **1998**, *12*, 27–35.
(17) Perdih, A.; Bren, U.; Solmajer, T. Binding free energy calculations of *N*-sulphonyl-glutamic acid inhibitors of MurD ligase. *J. Mol. Model.* **2009**, *15*, 983–996.
(18) Bren, U.; Martinek, V.; Florian, J. Free energy simulations of uncatalyzed DNA replication fidelity: structure and stability of T•G and dTTP•G terminal DNA mismatches flanked by a single dangling nucleotide. *J. Phys. Chem. B* **2006**, *110*, 10557–10566.
(19) Kamiya, N.; Yonezawa, Y.; Nakamura, H.; Higo, J. Protein−inhibitor flexible docking by a multicanonical sampling: Native complex structure with the lowest free energy and a free-energy barrier distinguishing the native complex from the others. *Proteins* **2008**, *70*, 41–53.
(20) Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. Flexible docking of a ligand peptide to a receptor protein by multicanonical molecular dynamics simulation. *Chem. Phys. Lett.* **1997**, *278*, 297–301.
(21) Berg, B. A.; Neuhaus, T. Multicanonical algorithms for first-order phase transitions. *Phys. Lett. B* **1991**, *267*, 249–253.
(22) Nakajima, N.; Nakamura, H.; Kidera, A. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J. Phys. Chem. B* **1997**, *101*, 817–824.
(23) Kim, J. G.; Fukunishi, Y.; Nakamura, H. Multicanonical molecular dynamics algorithm employing adaptive force-biased iteration scheme. *Phys. Rev. E* **2004**, *70*, 057103.
(24) Fukunishi, Y.; Mikami, Y.; Nakamura, H. The filling potential method: A method for estimating the free energy surface for protein-ligand docking. *J. Phys. Chem. B* **2003**, *107*, 13201–13210.
(25) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.
(26) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.
(27) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* **2002**, *296*, 1832–1835.
(28) Park, S.; Khalili-araghi, F.; Tajkhorshid, E.; Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **2003**, *119*, 3559–3566.
(29) Woo, H.; Roux, B. Calculation of absolute protein−ligand binding free energy from computer simulations. *Proc. Nat. Acad. Sci. U. S. A.* **2005**, *102*, 6825–6830.

(30) Frenkel, D.; Smit, B. Free energy calculations. In *Understanding Molecular Simulation from Algorithms to Applications*; Academic Press: San Diego, 1996; pp 152−156.
(31) Dang, L. X. Mechanism and thermodynamics of ion selectivity in aqueous solution of 18-crown-6 ether: A molecular dynamics study. *J. Am. Chem. Soc.* **1995**, *117*, 6954–6960.
(32) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E., Jr.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.
(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating lipid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
(35) Martyna, G. J.; Tuckerman, M E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996**, *87*, 1117–1157.
(36) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.
(37) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
(38) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an $N\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(39) Fukunishi, Y.; Suzuki, M. Potential of mean force calculation of solute molecules in water by a modified solvent accessible surface method. *J. Comput. Chem.* **1997**, *18*, 1656–1663.
(40) Kowall, T.; Geiger, A. Molecular dynamics simulation study of 18-crown-6 in aqueous solution. 2. Free energy profile for the association $18C6\cdots K^+$ in water. *J. Phys. Chem.* **1995**, *99*, 5240–5246.
(41) Bonas, G.; Bosso, C.; Vignon, M. R. Fast atom bombardment mass spectrometry for crown ether-alkali cation stability constant determination. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 88–89.
(42) Greengard, L.; Rokhlin, V. A fast algorithm for particle simulations. *J. Comput. Phys.* **1987**, *73*, 325–348.
(43) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. Direct calculation of the binding free energies of FKBP ligand. *J. Chem. Phys.* **2005**, *123*, 084108.
(44) Giannetti, A. M.; Koch, B. D.; Browner, M. F. Surface plasmon resonance based assay for the detection and characterization of promiscuous inhibitors. *J. Med. Chem.* **2008**, *51*, 574–580.
(45) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Overview of protein motions. In *Proteins*; Wiley: New York, 1988; pp 14−21.

CI9002156