Massively parallel computation of absolute binding free energy with well-equilibrated states

Hideaki Fujitani,* Yoshiaki Tanida, and Azuma Matsuura
Fujitsu Laboratories Ltd., 10-1 Morinosato-Wakamiya, Atsugi 243-0197, Japan
(Received 4 November 2008; published 26 February 2009)

A force field formulator for organic molecules (FF-FOM) was developed to assign bond, angle, and dihedral parameters to arbitrary organic molecules in a unified manner including proteins and nucleic acids. With the unified force field parametrization we performed massively parallel computations of absolute binding free energies for pharmaceutical target proteins and ligands. Compared with the previous calculation with the ff99 force field in the Amber simulation package (Amber99) and the ligand charges produced by the Austin Model 1 bond charge correction (AM1-BCC), the unified parametrization gave better absolute binding energies for the FK506 binding protein (FKBP) and ligand system. Our method is based on extensive work measurement between thermodynamic states to calculate the free energy difference and it is also the same as the traditional free energy perturbation. There are important requirements for accurate calculations. The first is a well-equilibrated bound structure including the conformational change of the protein induced by the binding of the ligand. The second requirement is the convergence of the work distribution with a sufficient number of trajectories and dense spacing of the coupling constant between the ligand and the rest of the system. Finally, the most important requirement is the force field parametrization.

DOI: 10.1103/PhysRevE.79.021914 PACS number(s): 87.15.−v, 05.10.−a

I. INTRODUCTION

The free energies of a molecular system describe its tendencies to associate and react, so they are desirable target quantities for the molecular computation. There have been various studies on the computation of free energies using rigorous statistical mechanical based simulations. Virtually all of the computations using explicit representation of water have taken advantage of the thermodynamic cycle to compute relative free energies between ligands or between mutant proteins. Free energy perturbation (FEP) and thermodynamic integration (TI) have been used to obtain the free energy difference between molecules ($\Delta\Delta G$) by the "mutations" of small fragments [1]. Molecular mechanics Poisson Boltzmann surface area (MM-PBSA) has been used to obtain absolute binding energies (ΔG) of large biomolecules and ligands [2]. MM-PBSA employs normal mode analysis to calculate the entropy from some snapshots of molecular dynamics simulations, but the value strongly depends on the snapshot. These methods were developed within the framework of the thermal equilibrium approach.

Jarzynski proved the nonequilibrium equality that the exponential average of nonequilibrium measurement of the work could yield the free energy difference between two thermodynamic states [3]. This was shown as a particular example of a more general nonequilibrium identity [4]. The Jarzynski identity was used to calculate free energies by "fast growth" nonequilibrium exponential work averaging [5,6], while the traditional "slow growth" performed reversible quasistatic switching of states. However, the results of exponential averaging strongly depend on the behavior at the tails of the distribution of work values, so it is difficult to derive the accurate free energy by exponential averaging. Shirts *et al.* demonstrated that the Bennett acceptance ratio method (BAR) can be interpreted in terms of the maximum likelihood estimate of the free energy difference given a set of nonequilibrium work values in the forward and reverse directions [7,8]. BAR yields lower variance for equal simulation time than exponential averaging or TI for many molecular simulations [9].

FKBP (FK506 binding protein) is best known as the target of the widely used immunosuppressive drug FK506 and is 107 residues long. The experimental binding affinities of the FKBP ligands were determined directly by the inhibition constant (K_i) of rotamase activity all in the same manner [10], making it a good test system for binding affinities. Pande's group calculated the binding energy of this system using the Folding@Home distributed computing [11]. Fujitani *et al.* also performed the same calculations on the Fujitsu BioServer massively parallel machine, using different force field parameters and different calculation procedures, which performed long thermal equilibration before the massively parallel free energy calculations [12]. This was named MP-CAFEE (massively parallel computation of absolute binding free energy with well-equilibrated states).

Fujitani *et al.* obtained binding free energies with root-mean-square deviation from a linear fit of only 0.4 kcal/mol and maximum deviation from the fit of 0.6 kcal/mol. However, the fitting line was shifted by 3.2 kcal/mol from the experimental values [12]. In order to explore the reason for this shift, Jayachandran *et al.* developed a new sampling method of parallelized-over-parts computation (POP), which used docking simulation and molecular dynamics [13]. Also there were other studies on it [14,15]. In this paper we explore the reason by applying different force field parametrization not only for the ligands but also the protein under the MP-CAFEE scheme. Developing our own tool for assigning force field parameters, we performed more comprehensive calculations for the binding energies of the FKBP ligand complexes.

*fjtani@labs.fujitsu.com

TABLE I. Bond and angle force field parameters of Amber99 and GAFF. K_r is in kcal/(mol Å^2), r_{eq} is in Å, K_θ is in kcal/(mol radian^2), and θ_{eq} is in degrees.

Amber99			GAFF		
Bond	K_r	r_{eq}	Bond	K_r	r_{eq}
C-C	310.0	1.525	c-c	290.1	1.550
CT-CT	310.0	1.526	c3-c3	303.1	1.535
C-CT	317.0	1.522	c-c3	328.3	1.508
C-CM	410.0	1.444	c-cc	377.4	1.462
CM-CM	549.0	1.350	cc-cd	504.0	1.371
C-N	490.0	1.335	c-n	478.2	1.345
C-O	570.0	1.229	c-o	648.0	1.214
C-O2	656.0	1.250	c-o	648.0	1.214
Angle	K_θ	θ_{eq}	Angle	K_θ	θ_{eq}
C-CT-H1	50.0	109.50	c-c3-h1	47.6	107.66
C-CT-HC	50.0	109.50	c-c3-hc	47.2	109.68
CA-CT-HC	50.0	109.50	ca-c3-hc	47.0	110.15
CC-CT-HC	50.0	109.50	cc-c3-hc	47.2	110.86
CT-CT-HC	50.0	109.50	c3-c3-hc	46.4	110.05
CT-CT-OH	50.0	109.50	c3-c3-oh	67.7	109.43
CT-CT-OS	50.0	109.50	c3-c3-os	67.8	108.42

II. FORCE FIELD

A. Amber model

In the 1990s, Kollman's group developed a second generation of the assisted model building with energy refinement (Amber) force field for the simulation of proteins, nucleic acids, and organic molecules in condensed phases [16]. In addition to improvements in the parameters, they tried to explicitly describe the algorithm by which the parameters were derived, so that consistent extensions could be made to molecules other than proteins. In the Amber force field, the total energy of the system is given by

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r_{ij} - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta_{ijk} - \theta_{eq})^2 + \sum_{\text{dihedrals}} \sum_{n} \frac{V_n}{2}[1 + \cos(n\phi_{ijkl} - \gamma)] + \sum_{i<j} \left\{ 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right\}. \quad (1)$$

Here, r_{eq} and θ_{eq} are equilibration structural parameters; K_r, K_θ, and V_n are force constants; n is multiplicity and γ is the phase angle for the torsional angle parameters. The ϵ_{ij} and σ_{ij} are van der Waals parameters and the partial charge q_i is assigned using the restrained electrostatic potential fit (RESP) model [17]. This "minimalist" representation of bond and angle energies is adequate for modeling most systems in the absence of mechanical strain like proteins in a thermal equilibrium. The goal of the force field is to accurately model conformational energies and intermolecular interactions involving proteins, nucleic acids, and other molecules with related functional groups which are of interest in organic and biological chemistry. In 2004, almost a decade later, Wang *et al.* developed a general Amber force field (GAFF) which provides parameters for most organic and pharmaceutical molecules composed of H, C, N, O, S, P, and halogens [18]. GAFF was designed to be consistent with traditional Amber force fields for proteins and nucleic acids.

Table I compares some bond and angle parameters of the same atom types between the ff99 force field in the Amber simulation package (Amber99) and GAFF. CT and c3 represent sp^3 carbon. CM, cc and cd represent sp^2 carbon in nonpure aromatic systems. There are significant differences in the bond force constant K_r. Amber99 assigns different bond parameters to carbonyl CO (C-O) and carboxylate COO^- (C-O2), but GAFF uses the same parameters. While GAFF assigns different angle parameters according to atom types, Amber99 uses the same parameters. GAFF parameters were much improved after the experience of the Amber99 force field parametrization. However, when we use Amber99 for a protein and GAFF for a ligand, the bond and angle parameters of the same atom types differ between the protein and ligand. We suspect this might introduce some errors into the molecular dynamics simulation, because the dynamics differ in the same bonds between the protein and ligand.

B. FF-FOM

The development of force field parameters is only part of the story. To correctly handle any molecule in molecular mechanics, one needs to automatically assign atom types and bond types, and then generate proper topologies that encode

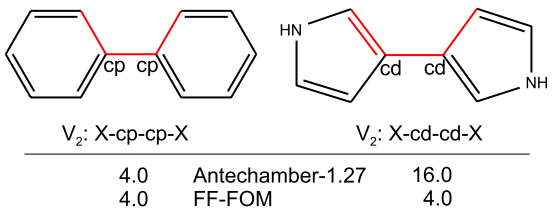

FIG. 1. (Color online) The V_2 torsional parameters assigned by Antechamber and FF-FOM in kcal/mol.

force field parameters for arbitrary organic molecules and their biological complexes. In order to assign the force field parameters to arbitrary organic molecules in a unified manner including proteins and nucleic acids, we developed our own force field assignment program named FF-FOM (force field formulator for organic molecules).

In order to identify the atom types and bond types of a target molecule, we first perform an empirical molecular orbital calculation with Austin Model 1 (AM1) to derive an optimized structure of the target molecule [19]. With the optimized structure FF-FOM first checks atom connectivity in the molecule using the covalent radius of the atoms and categorizes each atom into four types: Isolated atom, terminal atom, atom in a chain, or atom in a ring. Then, atom types and bond types are identified from the number of bonded atoms, intrinsic nature of atomic species, presence or absence of ring structure, net charge of a molecule, bond length, and so on. The current version of FF-FOM implements the GAFF model [18] and AM1-BCC (bond charge correction) charge model [20,21]. Since FF-FOM can incorporate user-defined parameters, it is easy to include RESP charges, which are calculated by *ab initio* molecular orbital programs and RESP utilities in the Antechamber program suite [22].

FF-FOM corrects some deficiencies of the present Antechamber implementation. One of them is assignment of torsion parameters between aromatic rings (Fig. 1). The V_2 torsion parameter is 14.5 kcal/mol for X-ca-ca-X dihedral where ca is an aromatic carbon and X is an arbitrary atom [16], but it is 4.0 kcal/mol for X-cp-cp-X dihedral where cp is a bridge aromatic carbon in biphenyl systems [18]. This difference comes from intraring torsion or interring torsion. However, since there is no bridge carbon type for bipyrrole systems in the current version of GAFF, Antechamber-1.27 assigns the V_2 parameter as 16.0 kcal/mol even for inter-ring X-cd-cd-X dihedral, where cd is an sp^2 carbon in nonpure aromatic systems such as a pyrrole molecule. The value of 16.0 kcal/mol is too large for the inter-ring torsion of bipyrrole. The value of V_2 is 4.0 kcal/mol for X-cp-cp-X dihedral and it is also 4.0 kcal/mol for X-ce-ce-X dihedral where ce is an inner sp^2 carbon in conjugated chain systems. Therefore, FF-FOM assigns 4.0 kcal/mol for the inter-ring X-cd-cd-X dihedral.

In order to determine GAFF parameters for all amino and nucleic acids in a unified manner, we first performed AM1 calculations for sample proteins and nucleic acids. Using the optimized structures, we assigned GAFF parameters for all amino and nucleic acids by FF-FOM. While Amber99 uses 13 atom types of carbon for proteins and nucleic acids, FF-FOM uses only 6 atom types of carbon (c, c2, c3, ca, cc, cd) from a total of 17 atom types of carbon in GAFF. The present version of FF-FOM generates topology files in two formats. One is for the GROMACS molecular dynamics suite [23–25] and the other is for the Amber program suite [26].

If Amber RESP charges are used instead of AM1-BCC charges for proteins and nucleic acids [27], GROMACS utilities can be used to generate topology files for proteins and nucleic acids instead of FF-FOM. The Amber force fields were ported to GROMACS by Pande's group [28]. We modified their Amber99 files to introduce the GAFF parametrization to GROMACS. Thus, GAFF topology files for proteins and nucleic acids can be generated by the GROMACS subprogram of pdb2gmx [29].

III. CALCULATION

A. Method

To calculate the binding free energy of a protein-ligand (PL) system, we perform two free energy calculations: One is for the solvated protein-ligand complex system and the other is for the solvated ligand system [30]. Keeping the native intramolecular interaction within the ligand, we annihilate the intermolecular interactions of the ligand to other molecules under an isothermal-isobaric condition.

$$\mathrm{PL}_{(\mathrm{SOL})} \rightarrow \mathrm{P}_{(\mathrm{SOL})} + \mathrm{L}_{(\mathrm{GAS})}, \quad \Delta G_{\mathrm{complex}}, \tag{2}$$

$$\mathrm{L}_{(\mathrm{SOL})} \rightarrow \mathrm{L}_{(\mathrm{GAS})}, \quad \Delta G_{\mathrm{sol}}, \tag{3}$$

$$\mathrm{PL}_{(\mathrm{SOL})} \rightarrow \mathrm{P}_{(\mathrm{SOL})} + \mathrm{L}_{(\mathrm{SOL})}, \quad \Delta G_{\mathrm{bind}} = \Delta G_{\mathrm{complex}} - \Delta G_{\mathrm{sol}}. \tag{4}$$

From the two kinds of calculations of Eqs. (2) and (3) we get the binding free energy ΔG_{bind} of Eq. (4).

The original Jarzynski identity was proved for the Helmholtz free energy of a constant *NVT* ensemble. Park and Schulten demonstrated that the Jarzynski identity holds for the Gibbs free energy in an isothermal-isobaric ensemble [31] and Cuendet derived it directly from the dynamics with the Nosé-Hoover thermostat [32].

The interaction strength of the ligand to other molecules is parametrized by a coupling constant λ in such a way that $\lambda_1 = 0$ means full interaction and $\lambda_n = 1$ means no interaction. The traditional slow growth FEP performs some amount of simulation steps at λ_i, then its final configuration is used as an initial configuration in the next simulation at λ_{i+1}. If a huge number of λ_i points were used, it might realize a reversible quasistatic process from the initial state at λ_1 to the final state at λ_n. If there is a hysteresis between forward and reverse results, the calculated free energy difference is not accurate. The reversibility is an important criteria to the free energy calculation within the framework of the thermal equilibrium approach.

There is another traditional FEP which does not pass the final configuration to the next simulation and just performs independent simulations at each λ_i. To calculate the free energy difference between λ_i and λ_{i+1} states, it takes an exponential average of the potential energy difference,

$$\Delta G_i = G_{i+1} - G_i = -RT \ln \left\langle \exp\left(-\frac{U_{i+1} - U_i}{RT} \right) \right\rangle_i, \quad (5)$$

where R is the gas constant and the average $\langle \cdots \rangle_i$ is taken in the trajectory at λ_i with the temperature T [1,33].

Now we consider nonequilibrium work measurement. We pick up one configuration from the λ_i state and measure the work W_i performed in changing the coupling constant value from λ_i to λ_{i+1} within a time τ. When the change is fast, the final λ_{i+1} state at τ may be unstable or far from equilibrium. But it does not matter. The Jarzynski identity requires thermal equilibrium of the initial state but it does not depend on how fast the work measurement is performed or what the final state is. In the zero limit of τ, W_i becomes $U_{i+1}-U_i$. Thus, the Jarzynski identify becomes Eq. (5). We can pick up many configurations from one equilibrium trajectory at λ_i to perform the work measurement. The Jarzynski identity gave a new point of view to the traditional FEP equation, which directly connects the work measurement with the free energy difference between two thermodynamic states. It tempted extensive work measurement by many simulations at each λ_i.

Before the free energy calculation we performed long equilibration for the solvated protein-ligand system and the solvated ligand system. Their final configurations were used as the initial configuration of all simulations at each λ_i, where we performed 12 simulations with different initial momentum distribution. We evaluated the free energy difference between two λ states by BAR using the work (potential difference) distribution in both directions.

Many previous works used auxiliary restraints to keep the ligand in the binding site during the absolute binding energy calculation. Since the protein-ligand interaction is weak near the end point ($\lambda=1$), the ligand moves out from the binding site without the restraint. When the ligand moves around the whole simulation system, it is difficult for the ligand to come back to the binding site when strengthening the protein-ligand interaction. A large hysteresis in G_{complex} between forward and reverse results near the end point was attributed to this irreversible process [34–36]. It was supposed that the auxiliary restraint made the ligand stay in the binding site and realized a reversible process so that the hysteresis disappeared. However, in a later section we show that the hysteresis near the end point does not appear even without any restraint or approximation and it is related with how to turn off the van der Waals interaction of the ligand.

Another reason for the auxiliary restraint was mentioned to give a relation to the standard concentration in experiment [36]. However, its method is based on the unreasonable energy shift of $RT \ln(VC^0)$ to a chemical potential, where V is the unit cell volume and C^0 is the unit concentration in molarity [Eq. (5) in Ref. [35]]. Contrary to the reasoning in Ref. [35] the energy shift term does not come out from the partition function for the Gibbs free energy. Thus, in reality, it does not gives any correction to the standard concentration with or without the restraint.

We calculate ΔG_{bind} by Eq. (4). If there is no interaction between the protein and ligand, $\Delta G_{\mathrm{complex}}$ is the same as ΔG_{sol}, and then $\Delta G_{\mathrm{bind}}=0$. Equation (4) gives a negative value for the attractive interaction and a positive value for the repulsive interaction between the protein and ligand. This is actually the definition of the absolute binding energy. Thus, no correction is needed to obtain the theoretical absolute binding energy. But there is a question about how to derive the experimental absolute binding energy, whose zero point exactly matches no interaction state between the protein and ligand.

The Gibbs free energy of a dilute solution is expressed in different ways depending on the units of concentration,

$$G = \Psi + RT \ln c_m = \Phi + RT \ln c_M, \quad (6)$$

where c_m and c_M are a solute concentration in mole fractions and in molarity, respectively. The standard Gibbs energy Ψ is related with the standard Gibbs energy Φ approximately by $\Psi=\Phi+RT \ln(1000\ \mathrm{cm}^3 \times \rho/M)$, where 1000 cm^3, M, and ρ are one liter, molecular weight, and density of water, respectively. The binding energy Ψ_{bind} in the mole fraction standard differs from the binding energy Φ_{bind} in the molarity standard

$$\begin{aligned} \Psi_{\mathrm{bind}} &= \Psi_{\mathrm{PL}} - \Psi_{\mathrm{P}} - \Psi_{\mathrm{L}} \\ &= \Phi_{\mathrm{PL}} - \Phi_{\mathrm{P}} - \Phi_{\mathrm{L}} - RT \ln(1000\ \mathrm{cm}^3 \times \rho/M) \\ &= \Phi_{\mathrm{bind}} - RT \ln(1000\ \mathrm{cm}^3 \times \rho/M) \end{aligned} \quad (7)$$

by $RT \ln(1000\ \mathrm{cm}^3 \times \rho/M)$, which is 2.4 kcal/mol at the room temperature. This difference does not depend on the solutes of protein and ligand. So it is a shift of the free energy zero point. In experiments the binding affinity is measured, for example, by a dissociation constant K_d which has the units of concentration. When K_d is in molarity, Φ_{bind} is given by $RT \ln K_d$. If K_d is in mole fractions, we get Ψ_{bind}. Φ_{bind} is closer to the absolute binding energy than Ψ_{bind}, but as far as we know there is no theoretical or experimental proof that Φ_{bind} meets the definition of the absolute binding energy. It might need a small constant shift to Φ_{bind} in order to obtain the experimental absolute binding energy. But more detailed conditions in experiments have bigger influence on Φ_{bind}. Therefore, we directly compare ΔG_{bind} with Φ_{bind}.

B. Molecular dynamics

The methodology was the same as the previous study of Fujitani *et al.* [12]. We used the same single precision version of GROMACS, which had two kinds of modifications from the original GROMACS-3.1.4. One was the annihilation of intermolecular interactions while keeping intramolecular interaction [37]. The other was to make some variables in the linear constraint solver (LINCS) to be double precision in order to improve energy conservation in constant-*NVE* calculations [38]. All molecular dynamics simulations were performed at 298 K with Nosé-Hoover temperature control [39,40] with a time constant of 0.3 ps and at 1.0 atm using Berendsen pressure control [41] with a time constant of 1.0 ps and a compressibility of 4.5×10^{-5}. We used the TIP3P water model, LINCS with order 8 to constrain all bonds, and 2 fs time step [42].

A neighbor list of 1.1 nm was utilized with an update frequency of 20 fs and van der Waals interactions were switched between 0.9 nm and 1.0 nm distances. Particle

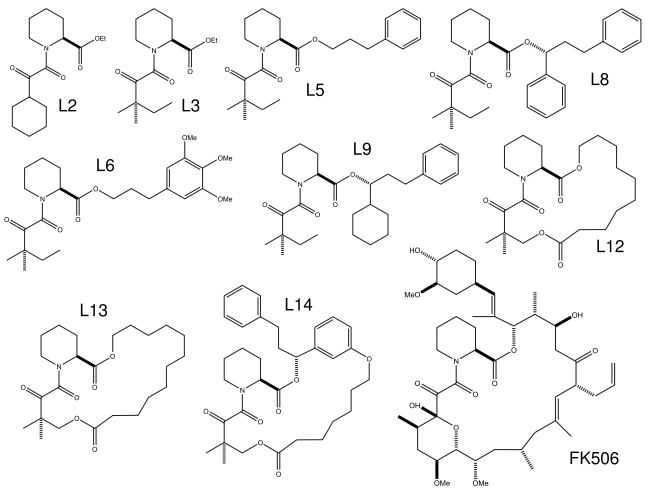

FIG. 2. Structures of 10 FKBP ligands.

mesh Ewald (PME) with an interpolation order of 4 was used for long-range Coulomb interactions beyond 1.1 nm. The Fourier spacing was approximately 0.12 nm. A long-range correction for the finite cutoff of the Lennard-Jones potential was taken into account for energy and pressure corrections [43].

All simulations were run in a truncated octahedron box. The solvated ligand system had 360–710 water molecules and the solvated FKBP-ligand complex system had 4800–5500 water molecules in the octahedron box. The complex system had six chlorine ions and two sodium ions to neutralize the charge of FKBP.

In the previous study of Fujitani *et al.* Amber99 parameters were used for the protein of FKBP and GAFF parameters for ligands with AM1-BCC charges. In this study we assigned GAFF parameters to FKBP and ligands in a unified manner by FF-FOM. Amber RESP charges were used for FKBP. RESP charges were calculated to ligands using the RESP utility of Antechamber after HF/6-31G* orbital calculations with GAUSSIAN98 [44]. As standard Amber calculations, the 1-4 interactions between hydrogen atoms were included.

We investigated 10 FKBP ligands in this study (Fig. 2), which is two more ligands (L13 and L14) than the previous study. In order to compare the results between different force field parameters, we also performed the calculations for L13 and L14 with the same parametrization as the previous study. As an initial configuration of the complex system, we used x-ray crystal structures for L8, L9, L13, and FK506 ligands protein data bank (Protein Data Bank code: 1fkg, 1fkh, 1fki, 1fkf) [10]. As all ligands have common binding elements, a pipecolate and an α-keto amide region, we assumed that bound conformations of the other ligands are similar to the known four complexes.

C. MP-CAFEE

After the insertion of water and ions around the solute, we performed energy minimization using the conjugate gradient method and a molecular dynamics simulation for 200 ps with the solute positions restrained. We then performed long equilibration at 298 K with the full interactions between the ligands and their surroundings, 5 ns for the solvated ligand system and from 10 ns to 50 ns for the solvated FKBP-ligand complex system. This long equilibration is important for the free energy calculation, because we should calculate the free energy of the stable thermodynamic state, not of transient, unstable thermodynamic states. The solvated ligand system easily reaches the equilibrated state with any initial momentum distribution of 298 K, but the solvated FKBP-ligand system sometimes suffered unstable dynamics and so it could not reach the stable bound conformation within our simulation time. Performing multiple long equilibration with different initial momentum distribution, we carefully checked whether the solvated FKBP-ligand system reached a stable bound conformation.

With the well equilibrated structure we performed massively parallel calculation of absolute binding free energy. Keeping the native intramolecular interaction within the ligand, the intermolecular interactions between the ligand and other molecules were annihilated using the soft-core potential [37] as shown by the equation:

$$U(\lambda^C,\lambda^{\mathrm{LJ}}) = \sum_{i,j}\left[(1-\lambda^C)\frac{q_i q_j}{r_{ij}} + 4(1-\lambda^{\mathrm{LJ}})\epsilon_{ij}\left(\frac{1}{[\alpha^{\mathrm{LJ}}\lambda^{\mathrm{LJ}}+(r_{ij}/\sigma_{ij})^6]^2} - \frac{1}{\alpha^{\mathrm{LJ}}\lambda^{\mathrm{LJ}}+(r_{ij}/\sigma_{ij})^6}\right)\right]. \quad (8)$$

We first turned off Coulomb interaction ($\lambda^C: 0 \rightarrow 1$) and then van der Waals interaction ($\lambda^{\mathrm{LJ}}: 0 \rightarrow 1$) with 0.5 for α^{LJ}. The λ spacing was chosen as the result became the same as that with closer λ^C and λ^{LJ} spacing [45]. Closer λ spacing is required to accurately calculate a free energy difference [46]. We used the 32 λ points: 11 λ^C (0.0, 0.1, 0.25, 0.4, 0.55, 0.65, 0.725, 0.8, 0.875, 0.95, and 1.0) and 21 λ^{LJ} (0.0, 0.1, 0.2, 0.3, 0.4, 0.475, 0.55, 0.6, 0.65, 0.675, 0.7, 0.725, 0.75, 0.775, 0.8, 0.825, 0.85, 0.875, 0.9, 0.95, and 1.0). For each λ state we performed 12 molecular dynamics simulations with different initial momentum distributions. Therefore we performed $32 \times 12 = 384$ molecular dynamics simulations to obtain one free energy.

When turning off the Coulomb interaction, full van der Waals interaction exists (λ^C, $\lambda^{\mathrm{LJ}}=0.0$). When turning off the van der Waals interaction, there is no Coulomb interaction ($\lambda^C=1.0$, λ^{LJ}). The 384 molecular dynamics simulations are performed independently, so there is no network communication between them. At each 100 fs time step we calculate potential energy differences between neighboring λ values (the work measurement). For example, in the simulation with $\lambda^C=0.8$, we calculate potential energy differences caused by 0.075 λ difference to the forward direction ($\lambda^C=0.875$) and to the reverse direction ($\lambda^C=0.725$).

The massively parallel calculations were performed on FUJITSU BioServer test machine, which has 1920 FR-V processors in a rack [47]. Each processor with eight way VLIW architecture has 256 MB memory and the Linux operating system, allowing us to run independent GROMACS

TABLE II. Typical simulation time (ns), CPU times (hours), and multiplicity of the simulation to calculate a binding free energy of a FKBP-ligand complex. The CPU time was measured by Pentium 4 (3 GHz).

	Equilibration			Free energy calculation			
System	MD	CPU	Multiplicity	MD	CPU	Multiplicity	Total CPU
Solvated ligand	5 ns	15 h	1	1 ns	3 h	384	1167 h (49 days)
Solvated complex	20 ns	600 h	3	2.5 ns	75 h	384	30600 h (1275 days)

molecular dynamics simulations. This arrangement is very suitable for performing the binding free energy calculations described here.

Table II shows typical simulation and computational times to calculate a binding energy for a FKBP-ligand system. For simplicity, the computational times on BioServer were converted to those on the popular CPU of Pentium 4 (3 GHz). MD means the simulation time and the multiplicity of 3 means three simulations are performed for the thermal equilibration. One of the three final configurations is used as the initial configuration for the massively parallel free energy calculation with the multiplicity of 384. Since the new version of GROMACS scales very well on parallel machines [48], all calculations can be parallelized efficiently. By a parallel machine with 300 CPUs, for example, a binding energy can be calculated in 4.5 days.

IV. RESULTS AND DISCUSSIONS

A. Well-equilibrated states

We check the interaction potential energy between the FKBP and ligand to find a stable binding conformation from multiple long equilibration of the solvated FKBP-ligand complex system. Figure 3 shows two cases of the long equilibration for the solvated FKBP-L2 system. Both cases started from the same initial configuration but with different momentum distribution of 298 K. The Coulomb interaction consists of short- and long-range contributions in the PME method. The upper line shows short-range Coulomb interaction and the lower line shows van der Waals interaction between the FKBP and L2.

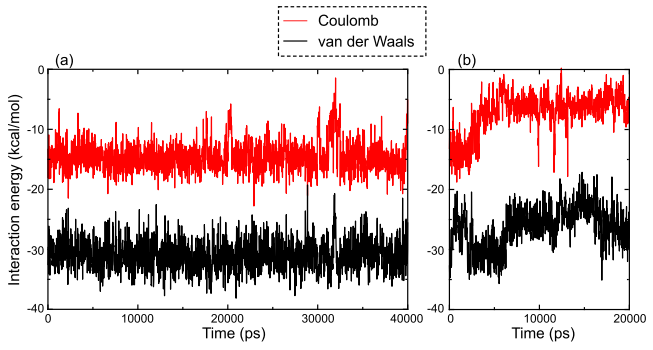

FIG. 3. (Color online) Potential energies versus simulation time for the solvated FKBP-L2 complex system. The upper line shows short-range Coulomb interaction and the lower line shows van der Waals interaction between FKBP and L2 ligand. Both cases of (a) and (b) started from the same initial configuration but with different momentum distribution of 298 K.

In the case (a) the system stays at the same energy level and occasionally shows larger fluctuations in the interaction energy. The pipecolate of L2 stayed atop of the indole of the 59th tryptophan (Trp-59) at the bottom of the hydrophobic pocket of FKBP during the 40 ns equilibration. It was a very stable trajectory. In the case (b) the Coulomb energy goes up and the van der Waals energy fluctuates, because L2 started to rotate in the hydrophobic pocket of FKBP. Even at 20 ns the pipecolate was around the middle of the pocket. It was still in a transient state. We use a well-equilibrated and stable conformation as the initial configuration for the massively parallel free energy calculation. The well-equilibrated state in the trajectory (a) is suitable for the initial configuration but the state in the trajectory (b) is not suitable for it.

Figure 4 shows the equilibration for the solvated FKBP-L14 system. L14 is the second largest ligand in this study and we did not have its x-ray crystal structure. During the equilibration the pipecolate of L14 always stayed at the bottom of the hydrophobic pocket and both interaction energies of short-range Coulomb and van der Waals changed to about 5 kcal/mol lower levels than the starting levels. It was caused by the conformational change of FKBP induced by the binding of L14. We must take account of this change to get an accurate binding energy.

Using a well-equilibrated configuration we performed 12 simulations at each λ point with different initial momentum distribution. There were 12 sets of the 32 λ point simulations. We call this set "sample." Figure 5 shows free energies of 100 ps versus the massively parallel simulation time for FKBP and L9 system. The free energy for a sample was

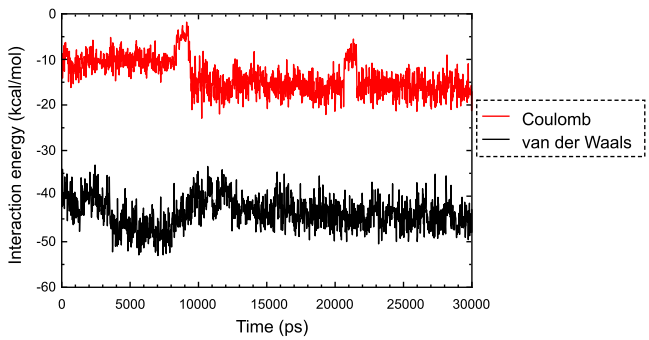

FIG. 4. (Color online) Potential energies versus simulation time for the solvated FKBP-L14 complex system. The upper line shows short-range Coulomb interaction and the lower line shows van der Waals interaction between FKBP and L14 ligand.

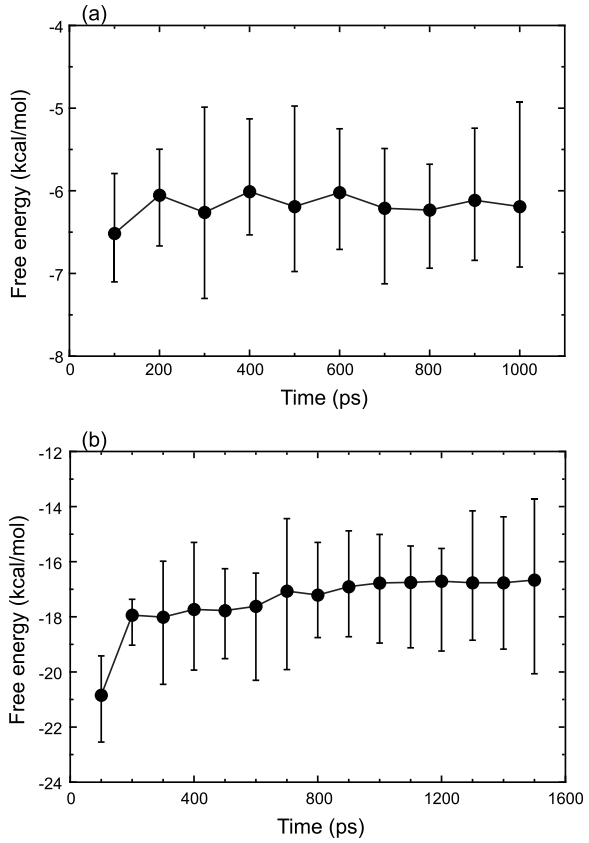

FIG. 5. Free energies versus simulation time in massively parallel calculation for FKBP and L9 ligand system. The upper figure (a) shows the free energy of the solvated L9 system, and the lower figure (b) shows the free energy of the solvated FKBP-L9 complex system. The free energy was calculated by BAR from 100 ps work distribution for each sample. The circle represents the average value among the 12 samples and the error bar indicates their minimum and maximum values.

calculated by BAR from its 100 ps work distribution. The circle shows the average of the 12 sample values and the error bar shows minimum and maximum values in the 12 samples. The upper figure (a) shows the free energy of the solvated L9 system and the lower figure (b) shows the free energy of the solvated FKBP-L9 complex system.

The 100 ps free energy largely changes at the beginning, because the well-equilibrated configuration with full interaction of the ligand to other molecules was used for all λ starting states of the massively parallel free energy calculation. Each state should go to equilibrium in accordance with the interaction strength of its λ value. To calculate the free energies of ΔG_{sol} and $\Delta G_{\mathrm{complex}}$ in Eq. (4) we use the work distribution in the equilibrated region in the massively parallel calculation. In the case of the FKBP and L9 system (Fig. 5) we used the region from 200 ps to 1000 ps in (a) for ΔG_{sol} and the region from 1000 ps to 15000 ps in (b) for $\Delta G_{\mathrm{complex}}$. Since the time to reach the equilibrium region depends on the system, we should carefully check the free energy profile like Fig. 5 in each case.

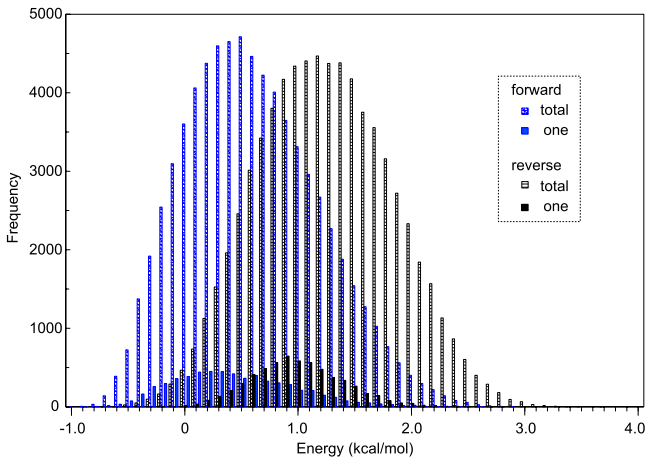

FIG. 6. (Color online) The work distribution between $\lambda^{\mathrm{LJ}}=0.775$ and $\lambda^{\mathrm{LJ}}=0.8$ with $\lambda^{C}=1$ obtained by the massively parallel calculation from 2000 ps to 2500 ps for the FKBP-L2 complex system. Measured works were counted by 0.1 kcal/mol spacing. The forward direction means to weaken the intermolecular interaction of L2 ligand and the reverse direction means to strengthen it. The larger histograms are the total work distribution of 12 trajectories and the smaller histograms are the work distribution obtained from one of the 12 trajectories.

B. Work distribution

Figure 6 shows an example of work distribution between two λ points ($\lambda^{\mathrm{LJ}}=0.775$ and $\lambda^{\mathrm{LJ}}=0.8$ with $\lambda^{C}=1$) for the solvated FKBP-L2 complex system. It was obtained from the equilibrated region from 2000 ps to 2500 ps of the massively parallel calculation. The forward direction means to weaken the intermolecular interaction of the ligand to other molecules and the reverse direction means to strengthen it. Since we measured the work at each 0.1 ps time step, there were 5000 measured works by a trajectory and 60 000 works in total by the 12 trajectories. The measured works were counted by 0.1 kcal/mol spacing in Fig. 6 and the smaller histograms are the work distribution obtained from one of the 12 trajectories.

The free energy difference between the λ points can be calculated by different methods. By the exponential averaging of Eq. (5) we obtained 0.79 kcal/mol from the forward work distribution, 0.97 kcal/mol from the reverse work distribution, and 0.87 kcal/mol by BAR from the work distribution in both directions. When we used the work distribution obtained only from one of the 12 trajectories (the smaller histograms), the values were 0.63 kcal/mol, 0.82 kcal/mol, and 0.66 kcal/mol, respectively. One molecular dynamics trajectory is not enough to get the work distribution in order to calculate the free energy difference. We need the total work distribution to accurately calculate free energy difference.

Each ΔG between the λ points is accumulated into a total ΔG of the system. Figure 7 shows λ dependence of the free energies calculated by exponential averaging of the forward and reverse work distributions for the FKBP and L2 system. The discrepancy between the forward and reverse directions is small and it also appears on the annihilation of the Cou-

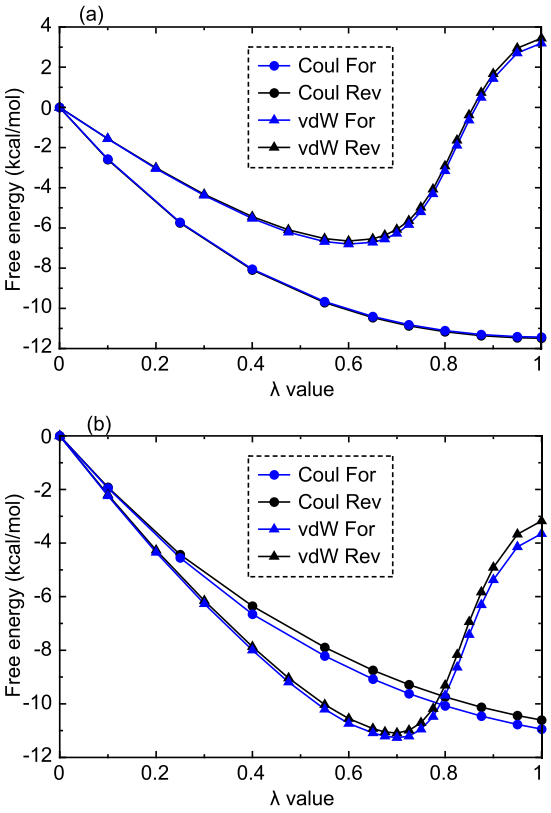

FIG. 7. (Color online) Free energies calculated by exponential averaging of the work distribution in both directions for the solvated L2 system (a) and the solvated FKBP-L2 complex system (b). The symbols show the 32 λ points.

lomb interaction in the complex system (b). Since the results of exponential averaging strongly depend on the behavior at the tails of the work distribution, it is difficult to get very accurate free energy by exponential averaging. Thus the discrepancy between the forward and reverse directions remains. To overcome the discrepancy we used BAR to calculate the free energy from the work distribution in both directions. It gives a line in the middle of the forward and reverse lines in Fig. 7.

We first annihilate the Coulomb interaction of the ligand and then van der Waals interaction. During the annihilation of Coulomb interaction the ligand keeps the van der Waals interaction to other molecules and almost stays around the binding site in the solvated complex system. When the van der Waals interaction becomes weaker, the ligand gets out from the binding site and moves around in the system. The whole Coulomb part and the left half of the van der Waals part in Fig. 7 make a favorable contribution to the solvation or binding energy and the right half of the van der Waals part makes an unfavorable contribution. For the 10 FKBP ligands the total free energies of van der Waals interaction in the solvated ligand system were positive values like (a) in Fig. 7.

In our calculation the discrepancy between the forward and reverse directions was small. The ΔG between $\lambda^{\mathrm{LJ}}=0.95$ and 1.0 in the solvated FKBP-L2 complex system was 0.49 kcal/mol from the forward direction and 0.50 kcal/mol from the reverse direction. Its discrepancy was only 0.01 kcal/mol. If we did not use the soft-core potential and

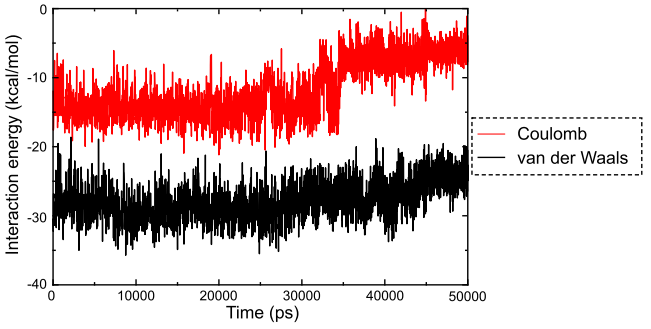

FIG. 8. (Color online) Potential energies versus simulation time for the solvated FKBP-L3 complex system. The upper line shows short-range Coulomb interaction and the lower line shows van der Waals interaction between FKBP and L3 ligand.

the auxiliary restraint, we obtained 2.77 kcal/mol from the forward direction and could not get a finite value from the reverse direction. Because at $\lambda^{\mathrm{LJ}}=1$ the ligand penetrates into other molecules and large work values by the van der Waals potential of other molecules give an infinite number by exponential averaging. The auxiliary restraint with some type of soft-core potential might be a possible method to get finite values without the hysteresis. But we prefer to use appropriate soft-core parameters to perform the free energy calculation without any approximation or restraint.

C. Rotation in the pocket

A bound conformation of the protein-ligand complex system is one of the key points for accurate calculation of binding free energy. Long equilibration yields a correct bound structure, including relaxation of water molecules and conformational change of the protein induced by the binding of the ligand. For the eight larger FKBP ligands (L5, L8, L6, L9, L12, L13, L14, FK506) we easily got stable equilibrated structures, which showed lower interaction energies between the FKBP and ligand. The pipecolate of the ligands always stayed at the bottom of the hydrophobic pocket of FKBP. However, for the small ligands of L2 and L3 the situation was slightly different. We performed several sets of long equilibration, changing not only initial momentum distribution but also initial conformation in the hydrophobic pocket in order to get the most stable bound structure.

As explained in the preceding section L2 started to rotate in the hydrophobic pocket even from the lowest interaction energy levels. Figure 8 shows another interesting trajectory of the solvated FKBP-L3 complex system. L3 stayed in the lowest energy state until 30 ns with the pipecolate at the bottom of the hydrophobic pocket. At 43 ns L3 rotated and the pipecolate moved up to the top of the pocket as shown in Fig. 9 which was drawn by the molecular visualization software of VMD [49]. It means that the rotation of the small ligands occurs even from the well-equilibrated state. In other trajectories the pipecolate rotated from the top to the bottom in the pocket. These rotations were common to the L2 and L3 complex systems, but they were never observed in other complex systems of larger ligands within our simulation time.

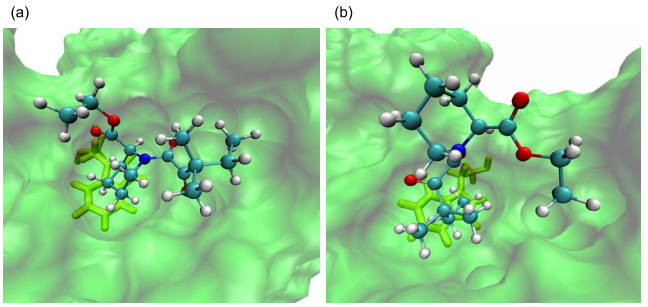

FIG. 9. (Color online) Bound structures of the FKBP-L3 complex in the trajectory of Fig. 8. (a) The lowest energy structure with the pipecolate at the bottom and (b) rotated structure with the pipecolate at the top in the hydrophobic pocket. Trp-59 at the bottom of the hydrophobic pocket is transparently shown.

In the previous study by the Amber99 and AM1-BCC model the conformation with the pipecolate at the top was more stable than the conformation with the pipecolate at the bottom for the L2 and L3 complex systems [12]. But in this study we obtained the most stable conformations of L2 and L3 complex with the pipecolate at the bottom of the pocket.

We checked whether these conformations had lower energies than other conformations obtained by a different method. Jayachandran *et al.* generated binding poses by the docking software Surflex [50] and performed POP computations [13]. We compared potential energies with their higher score's bound poses of L2 and L3. Since our force field parameters were applied to their docking poses, we first performed energy minimization for their poses by the conjugate gradient method and then calculated the interaction energies (Fig. 10). The abscissa means different poses. The number zero is our bound structures which were used in the free energy calculation. The docking pose had various conformations with the pipecolate at the bottom, middle, or top of the pocket. Among them the number zero had the lowest interaction energy. Since we compared the interaction energy between various conformations of the same ligand in the same hydrophobic pocket, the entropy contribution to the free energy should be almost the same. So the number zero should give the lowest binding free energy among the binding poses in Fig. 10.

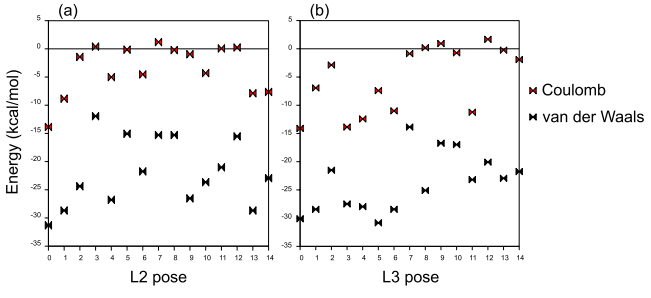

FIG. 10. (Color online) Bound conformations versus interaction energies between the FKBP and L2 or L3 ligand. The upper signs mean short-range Coulomb interaction energies and the lower signs mean van der Waals interactions. The number zero is our stable bound structure obtained by the molecular dynamics equilibration and others are docked poses obtained by Jayachandran *et al.* [13].

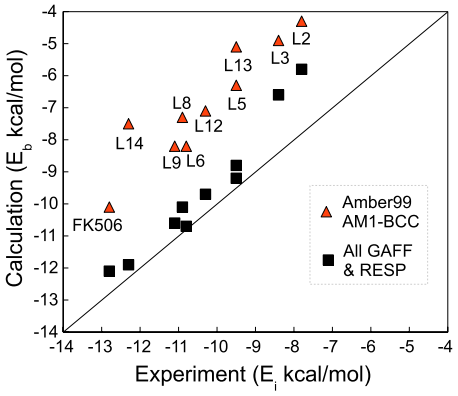

FIG. 11. (Color online) Two kinds of calculated absolute binding free energies versus experimental binding energies obtained from inhibition constants for the 10 FKBP ligands.

D. Binding energy

Our absolute binding free energies calculated by BAR are compared with the experimental values obtained from inhibition constants K_i of Holt *et al.* by $RT \ln K_i$ [10], where R is the gas constant and $T=298$. In Fig. 11 the abscissa and ordinate have the same energy range so the diagonal line means perfect agreement between calculation and experiment. The triangles show binding energies calculated with Amber99 for FKBP and AM1-BCC charges for the ligands. The absolute binding energies of the eight ligands in the previous study showed good linearity but were shifted by 3.2 kcal/mol on average from experimental values [12]. New calculated values of L13 and L14 with the same Amber99 and AM1-BCC model deviate from the previous line of the eight ligands. The rectangles show new calculated values with the unified force field model of GAFF parameters and RESP charges for the FKBP and ligands. They are closer to the diagonal line than the triangles. Especially, the eight larger ligands have better agreement within 1 kcal/mol difference from experimental values.

Table III lists ligand solvation energies ΔG_{sol}, complex free energies $\Delta G_{\mathrm{complex}}$, and binding energies ΔG_{bind} for two force field models. The values in parentheses are root-mean-square deviation within the 12 samples in the massively parallel calculation. As the same GAFF bond parameters and van der Waals parameters were used for the ligands in the two models, ΔG_{sol} differences between the two models came only from the difference of the ligand charge model between AM1-BCC and RESP. $\Delta G_{\mathrm{complex}}$ differences depend on not only the ligand charge model but also the bond parameter difference of the protein between Amber99 and GAFF. The unified force field model gave better agreement of ΔG_{bind} with the experimental value ΔG_{exp}. Both $\Delta G_{\mathrm{complex}}$ and ΔG_{sol} contribute to this improvement, depending on the ligand.

AM1-BCC was designed to yield charge sets of comparable quality to RESP charges [20]. The AM1-BCC ΔG_{sol} shows the same trend as the RESP ΔG_{sol}, but the former shows stronger solvability than the latter for all ligands. The former values of L13 and L14 are larger than the latter values by more than 3 kcal/mol. This is why L13 and L14 deviated from the previous line of the eight ligands with the Amber99 and AM1-BCC model in Fig. 11.

TABLE III. Comparison of calculated free energies for the FKBP and ligands between two kinds of force field models. The size means number of atoms of each ligand. ΔG_{exp} is the experimental value obtained by $RT \ln K_i$. $\Delta G_{bind}=\Delta G_{complex}-\Delta G_{sol}$, where ΔG_{sol} is the free energy calculated in the solvated ligand system and $\Delta G_{complex}$ is the free energy calculated in the solvated FKBP-ligand complex system. The values in parentheses are root-mean-square deviation within the 12 samples in the massively parallel calculation. All values are in kcal/mol. ΔG_{bind} versus ΔG_{exp} was plotted in Fig. 11.

			Amber99+AM1BCC			GAFF+RESP		
Ligand	Size	ΔG_{exp}	ΔG_{sol}	$\Delta G_{complex}$	ΔG_{bind}	ΔG_{sol}	$\Delta G_{complex}$	ΔG_{bind}
L2	46	−7.8	−8.5	−12.8	−4.3	−8.1 (0.1)	−13.9 (0.6)	−5.8
L3	45	−8.4	−7.6	−12.5	−4.9	−6.9 (0.1)	−13.5 (0.6)	−6.6
L5	58	−9.5	−8.4	−14.7	−6.3	−7.6 (0.1)	−16.4 (0.9)	−8.8
L6	70	−10.8	−12.7	−20.9	−8.2	−10.4 (0.4)	−21.1 (1.2)	−10.7
L8	68	−10.9	−9.3	−16.6	−7.3	−7.5 (0.1)	−17.6 (0.7)	−10.1
L9	74	−11.1	−7.1	−15.3	−8.2	−6.2 (0.2)	−16.8 (0.7)	−10.6
L12	64	−10.3	−15.4	−22.5	−7.1	−12.8 (0.1)	−22.5 (0.6)	−9.7
L13	70	−9.5	−14.6	−19.7	−5.1	−11.6 (0.2)	−20.8 (0.7)	−9.2
L14	85	−12.3	−16.5	−24.0	−7.5	−12.7 (0.2)	−24.6 (1.1)	−11.9
FK506	126	−12.8	−26.0	−36.1	−10.1	−24.6 (0.5)	−36.7 (0.9)	−12.1

Figure 12 shows the atomic charges of L14 for the two charge models. The AM1-BCC model well reproduces the trend of the RESP model even for L14, which shows the biggest difference of ΔG_{sol} between the two models. Since the absolute binding free energy calculation needs highly quantitative accuracy, the present version of the AM1-BCC model is not sufficiently accurate for the free energy calculation.

As discussed in the preceding subsection we used the binding poses of L2 and L3 with the lowest energy among the examined poses as the starting structure of the massively parallel calculation of the absolute binding free energy. However, the binding free energies of L2 and L3 slightly deviated from the linearity of other larger ligands (Fig. 11). The calculation showed slightly weaker binding energy than the experiment. We suspect this might be related with the rotation of L2 and L3 in the FKBP binding pocket, which would disturb the ligand to diffuse out from the pocket. Then, slightly stronger inhibition was observed in the experiment. Of course this conjecture should be verified by other experiment.

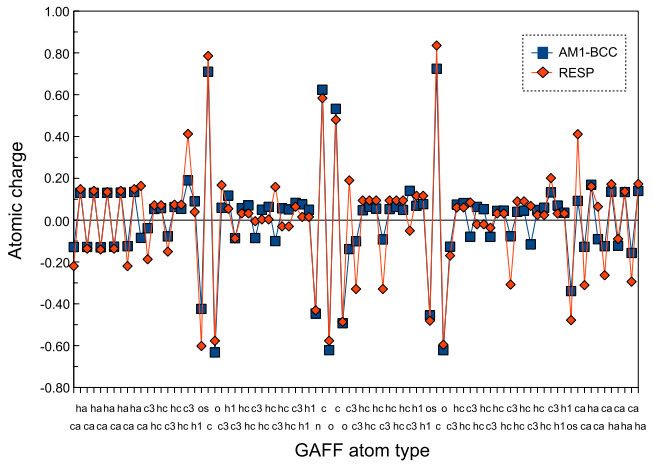

FIG. 12. (Color online) Atomic charges of L14 ligand. The rectangle shows the AM1-BCC charge and the diamond shows the RESP charge on each atom.

V. CONCLUSIONS

The nonequilibrium equality for the free energy difference between thermodynamic states gives concrete basis and efficient framework for the massively parallel computation of the absolute binding free energy for biomolecules. Thus, we can utilize a large number of computer processors.

There are some important requirements for actual calculations to get accurate absolute binding energies. The first that we require is a well-equilibrated bound structure including the conformational change of the protein induced by the binding of the ligand. The second requirement is convergence of the work distribution including a sufficient number of λ points and trajectories. Finally the force field parametrization is the most important issue.

The GAFF parameters and RESP charges were assigned to the protein and ligand in a unified manner. In comparison with the previous model of Amber99 and AM1-BCC the unified model gave better absolute binding energies. This implies that MP-CAFEE may help to develop the force field parameters for biomolecules, because one of the major obstacles is the difficulty of justifying the parametrization by quantitative comparison with experimental data. In our experience the absolute binding energy is sensitive to the force field parameter of not only ligands but also proteins and nucleic acids depending on their binding structure. The improved force field parameters would give better binding energies.

We think GAFF parametrization was much improved from Amber99, but the dihedral torsional parameter should

be improved. Pérez *et al.* performed high level *ab initio* calculations to improve the Amber dihedral parameters in nucleic acids [51]. They verified their new dihedral parameters by performing a very extensive comparison between molecular dynamics simulations and experimental data. Their improvement is also applicable to the GAFF parametrization, because GAFF has the same related dihedral parameters as Amber99. In order to prove the real power of the "minimalist" model of Eq. (1), much work remains to be done.

From the comparison of bound structures, we conjectured that the rotation of L2 and L3 within the hydrophobic pocket might disturb the ligand to diffuse out. Then, slightly stronger inhibition was observed in the experiment. Our conjecture should be verified by other experiment. Nevertheless, it is clear that more accurate and more extensive molecular dynamics simulations will elucidate many hidden phenomena in biochemistry.

The development of computer technology is another aspect of the simulation. Without BioServer it would have been difficult to develop our methodology. Shaw's group designed a new special purpose machine to bring millisecond-scale molecular dynamics simulations within reach for molecular systems involving tens of thousands of atoms [52]. Millisecond-scale simulations will clarify the folding process of pharmaceutical target proteins. With more accurate force field parametrization advanced computation will open up a new landscape in biochemistry.

ACKNOWLEDGMENTS

We would like to thank Vijay Pande (Stanford University) and all his group members, especially Guha Jayachandran, Michael R. Shirts, Christopher D. Snow, and Eric J. Sorin for useful discussions and giving us their tools and docking poses, and Erik Lindahl (Stockholm University) for help in modifying GROMACS. This work was partly supported by the High-Throughput Biomolecule Analysis System Project of NEDO (New Energy and Industrial Technology Development Organization, Japan). We would like to thank all members of the BioServer project in Fujitsu.

[1] P. A. Kollman, Chem. Rev. (Washington, D.C.) **93**, 2395 (1993).
[2] J. Srinivasan, T. E. Cheatham, P. A. Kollman, and D. A. Case, J. Am. Chem. Soc. **120**, 9401 (1998).
[3] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[4] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[5] D. A. Hendrix and C. Jarzynski, J. Chem. Phys. **114**, 5974 (2001).
[6] G. Hummer, J. Chem. Phys. **114**, 7330 (2001).
[7] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).
[8] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[9] M. R. Shirts and V. S. Pande, J. Chem. Phys. **122**, 144107 (2005).
[10] D. A. Holt, J. I. Luengo, D. S. Yamashita, H. J. Oh, A. L. Konialian, H. K. Yen, L. W. Rozamus, M. Brandt, M. J. Bossard, M. A. Levy, D. S. Eggleston, J. Liang, L. W. Schultz, T. J. Stout, and J. Clardyt, J. Am. Chem. Soc. **115**, 9925 (1993).
[11] M. R. Shirts, Ph.D. thesis, Department of Chemistry, Stanford University, 2004.
[12] H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin, and V. S. Pande, J. Chem. Phys. **123**, 084108 (2005).
[13] G. Jayachandran, M. R. Shirts, S. Park, and V. S. Pande, J. Chem. Phys. **125**, 84901 (2006).
[14] J. Wang, Y. Deng, and B. Roux, Biophys. J. **91**, 2798 (2006).
[15] M. R. Shirts, D. L. Mobley, J. D. Chodera, and V. S. Pande, J. Phys. Chem. B **111**, 13052 (2007).
[16] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[17] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, J. Phys. Chem. **97**, 10269 (1993).
[18] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).
[19] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902 (1985).
[20] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **21**, 132 (2000).
[21] A. Jakalian, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **23**, 1623 (2002).
[22] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, J. Mol. Graphics Modell. **25**, 247 (2006).
[23] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[24] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).
[25] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, J. Comput. Chem. **26**, 1701 (2005).
[26] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. Woods, J. Comput. Chem. **26**, 1668 (2005).
[27] P. Cieplak, W. D. Cornell, C. Bayly, and P. A. Kollman, J. Comput. Chem. **16**, 1357 (1995).
[28] E. J. Sorin and V. S. Pande, Biophys. J. **88**, 2472 (2005).
[29] See EPAPS Document No. E-PLEEE8-79-115902 for the unified force field files for pdb2gmx. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.
[30] W. L. Jorgensen, J. K. Buckner, S. Boudon, and J. Tirado-Rives, J. Chem. Phys. **89**, 3742 (1988).
[31] S. Park and K. Schulten, J. Chem. Phys. **120**, 5946 (2004).
[32] M. A. Cuendet, Phys. Rev. Lett. **96**, 120602 (2006).
[33] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[34] B. Roux, M. Nina, R. Pomes, and J. C. Smith, Biophys. J. **71**, 670 (1996).

[35] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, Biophys. J. **72**, 1047 (1997).
[36] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus, J. Phys. Chem. B **107**, 9535 (2003).
[37] M. R. Shirts and V. S. Pande, J. Chem. Phys. **122**, 134508 (2005).
[38] R. A. Lippert, K. J. Bowers, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, and D. E. Shaw, J. Chem. Phys. **126**, 046101 (2007).
[39] S. Nosé, Mol. Phys. **52**, 255 (1984).
[40] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[41] H. J. C. Berendsen, J. P. Postma, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[42] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).
[43] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).
[44] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, and J. A. Pople, GAUSSIAN98, Revision A.7, Gaussian, Inc., Pittsburgh, PA, 1998.
[45] Y. Tanida, M. Ito, and H. Fujitani, Chem. Phys. **337**, 135 (2007).
[46] D. A. Kofke, Fluid Phase Equilib. **228-229**, 41 (2005).
[47] H. Okano, A. Suga, T. Shiota, Y. Takebe, Y. Nakamura, N. Higaki, H. Kimura, H. Miyake, T. Satoh, K. Kawasaki, R. Sasagawa, W. Shibamoto, M. Sasaki, N. Ando, T. Yamana, I. Fukushi, S. Tago, F. Hayakawa, T. Kamigata, S. Imai, A. Satoh, Y. Hatta, N. Nishimura, Y. Asada, T. Satoh, T. Sukemura, S. Ando, and H. Takahashi, ISSCC Digest Technical Papers, 2002, pp. 374–375.
[48] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).
[49] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[50] A. N. Jain, J. Med. Chem. **46**, 499 (2003).
[51] A. Pérez, I. Marchá, D. Svozil, J. Sponer, T. E. Cheatham III, C. A. Laughton, and M. Orozco, Biophys. J. **92**, 3817 (2007).
[52] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossváry, J. L. Klepeis, T. Layman, C. McLeavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang, *Anton: A Special-Purpose Machine for Molecular Dynamics Simulation*, in Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA '07), San Diego, California, June 9–13, 2007, pp. 1–12.