Ranking Ligand Binding Affinities With Avidin: A Molecular Dynamics-Based Interaction Energy Study

Jian Wang, Richard Dixon, and Peter A. Kollman*
Department of Pharmaceutical Chemistry, University of California, San Francisco, California

ABSTRACT **The binding of 14 biotin analogues to avidin is examined to evaluate the viability of calculating binding free energy based on molecular dynamics (MD) trajectories. Two approaches were investigated in this work. The first one uses the linear interaction energy approximation, while the other approach utilizes the interaction free energy. The results obtained from these two methods were found to correlate well with the experimental binding free energy data for 10 out of 14 ligands. For the other four ligands, both methods overestimate their binding strength by more than 7 kcal/mol. Free energy calculations using the thermodynamic integration method are employed to understand this overestimation. The effect of protein flexibility on binding free energy calculation and the effect of charged or neutral ligands on the calculated results are discussed. MD simulations are shown to be able to provide insight into the interactions occurring in the active site and the origins of variations in binding free energy. Proteins 1999;34:69–81.** © 1999 Wiley-Liss, Inc.

Key words: biotin; molecular dynamics; free energy calculations; docking

INTRODUCTION

Given a receptor structure complexed with a ligand, a common strategy for structure-based drug design is to search for optimal binding modes of a candidate ligand at the binding site, and then to evaluate the ligand binding affinity in order to provide mechanistic insight into the ligand-receptor interaction or to predict binding affinity of a ligand prior to organic synthesis. The search for optimal binding modes is a docking problem.[1,2] Many algorithms have been developed to effectively explore the energy landscape of a ligand at the binding site. Ranking ligand binding affinities, however, presents an additional large challenge in molecular modeling. Several empirical scoring functions have been developed in this regard and have been described by Böhm,[3] Williams,[4] Marshall,[5] Murray,[6] and others. The basic idea of this approach is to derive an empirical function composed of free energy components, using the known ligand-protein complex structures and their experimental binding free energies. Examples of such free energy components are hydrogen bonding energy, lipophilic energy, loss of translational and rotational entropy and others. This approach is computationally efficient, but its application is often limited. Since the number of experimentally known ligand-receptor complex structures is still small, the basis set for deriving empirical parameters of an empirical scoring function (i.e., to determine the relative weighting factors among free energy components) is far from complete. Furthermore, the scoring function is derived from the known ligand-receptor structures, where receptor binding sites have perfect match with the corresponding ligands. As a result, the scoring function works best only for well matched ligand-receptor complex structures, but it may have little tolerance to a slight change in binding orientation or conformation of a ligand as well as not be able to correctly penalize suboptimal structures. For a set of ligands with different sizes and heteroatoms, it is often difficult to obtain a perfectly correct binding structure using current state-of-art docking programs;[7–10] thus, the subsequent prediction of binding affinity using an empirical scoring function is often questionable.

The free energy calculation methods including thermodynamic integration (TI) and free energy perturbation (FEP) methods, on the other hand, provides a rigorous theoretical framework to accurately estimate relative binding free energy of ligands at binding site. In practice, free energy calculations often run into sampling problem (both in conformation and configuration sampling), because free energy calculations spend most of their computation time on non-physical intermediate states. This problem becomes even more severe when TI or FEP are applied to predict binding affinities of a set of ligands with very dissimilar structural scaffolds. In this case, a very long simulation, which is at least a few orders of magnitude slower than the empirical scoring approach, is required in order to even hope to accurately calculate relative free energies of binding.

Recently, Åqvist et al. have proposed a semiempirical linear interaction energy (LIE) approach for estimating absolute ligand binding free energies.[14–17] In this approach, the binding free energy is approximated as

$$\Delta G_{bind} = \alpha(< E^{ES} >_{pro} - < E^{ES} >_{aq}) + \beta(< E^{VW} >_{pro} - < E^{VW} >_{aq}) \quad (1)$$

where E^{ES} and E^{VW} are the electrostatic and van der Waals interaction energies between ligand and its surroundings

Grant sponsor: National Institutes of Health; Grant number RR-1081.
*Correspondence to: Peter A. Kollman, Department of Pharmaceutical Chemistry University of California, San Francisco, CA 94143–0446. E-mail: pak@cgl.ucsf.edu
Received 17 April 1998; Accepted 27 August 1998

in protein (*pro*) or in aqueous solution (*aq*), respectively; $< >$ denotes the ensemble average over molecular dynamics or Monte Carlo simulations trajectory; and α and β are empirical parameters. In several test cases, Åqvist et al. have found that $\alpha \approx 0.5$ and $\beta = 0.16$ are transferable among different protein systems. When using LIE to calculate the binding free energies for P450cam-substrate complexes, Paulsen and Ornstein have found that $\alpha = 0.5$ and $\beta = 1.043$ are best to correlate the calculated binding free energies with the respective experimental values.[18] Jones-Herzog and Jorgensen have also observed that $\alpha = 0.5$ and $\beta = 0.161$ are not optimal for sulfonamide inhibitors with human thrombin.[19] It seems that these two parameters can be protein-dependent. Regardless of the transferability of these parameters, there are several appealing features of the LIE approach. First of all, LIE is faster than FEP calculations (more than a factor of five) because LIE directly simulates the end states. Secondly, LIE takes into account the protein flexibility just as FEP does, which is often not included in scoring function approaches. Finally, LIE uses an explicit solvent model in its simulation; thus, the desolvation free energy can be reasonably handled.

These advantages make LIE a potentially useful tool in structure-based lead optimization. During the lead optimization stage, one often needs a relatively reliable computational method to understand detailed interactions between the lead compounds and their receptor, and more importantly, to estimate binding affinities for a set of suggested compounds prior to synthesis. The aim of this paper is to apply the LIE method to calculate ligand binding free energies for biotin analogues interacting with avidin.

Enzyme avidin is a basic tetrameric glycoprotein (4×128 amino acid residues) isolated from hen egg-white.[20] Each of its protomers is organized in an eight-stranded antiparallel orthogonal β-barrel, with extended loop regions. Avidin binds to vitamin biotin with exceptionally high affinity constants ($K_d \sim 10^{-15}$ M).[21,22] This high binding affinity suggests that avidin is probably an antibiotic protein inhibiting bacterial growth and has lead to many attempts in the biotechnology industry to utilize this unique activity.[23]

The avidin binding site, which is located at the center of the barrel, consists of both hydrophobic and polar residues for recognition of the tightly bound vitamin. The binding of biotin to avidin involves a number of hydrogen bonds (H-bonds) and van der Waals (vdW) interactions as revealed by crystal structures.[24,25] In the free energy perturbation study of biotin binding to streptavidin, a structural homologue of avidin, Miyamoto and Kollman observed that the vdW energy is significantly larger than the electrostatic and hydrogen bonding energies in the binding of biotin by streptavidin.[26,27] The binding site of streptavidin and avidin have a nearly ideal complementary shape for biotin, the binding of biotin to streptavidin does not lead to any significant reorganization of binding site with the exception of one mobile loop. There is little free energy price to pay for streptavidin to create a cavity upon biotin binding. This study provides a good starting point for us. Furthermore, many biotin analogues as well as a variety of molecules with no apparent structural similarity to biotin have been reported to bind to avidin.[21] The range of their binding free energies is over 15 kcal/mol, larger than almost any comparable system. So avidin is an ideal choice to test the LIE methods and to further understand the high binding affinity of biotin for avidin.

METHODS AND COMPUTATIONAL DETAILS

Force Field and Molecular Dynamics Simulations

All simulations presented in this paper were carried out using a modified molecular simulation package Amber4.1.[28] The pairwise potential energy function of the force field has the form:

$$V = \sum_{bonds} k_r(r - r_{eq})^2 + \sum_{\substack{bond \\ angles}} k_\theta(\theta - \theta_{eq})^2 + \sum_{torsions} \frac{V_n}{2}[1 + \cos(n\varphi - \varphi_o)] + \sum_{i<j} \epsilon_{ij}\left[\left(\frac{R^*_{ij}}{r_{ij}}\right)^{12} - \left(\frac{R^*_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i<j} \frac{q_i\, q_j}{r_{ij}} \qquad (2)$$

where k_r, k_θ, V_n, r_{eq}, θ_{eq}, and φ_0 are empirical parameters relating to bond, bond angles and torsion angles, ϵ_{ij} and R^*_{ij} are van der Waals parameters, and q_i are the atomic charges, respectively. The atomic charges (q_i) used in this work were derived from the electrostatic potentials calculated at the HF/6–31G* level with AM1 optimized geometry.

In order to calculate absolute binding free energies, two molecular dynamics simulations were carried out for each ligand: one for unbound ligands surrounded by an 18 Å sphere of waters and the other for ligand-avidin complexes together with a cap of waters outside the binding pocket. The cap of waters in the complexes were added to fill space up to 18 Å from the center of mass of the ligand. Since the crystal structure is available only for biotin-avidin (pdb entry: 1avd), the initial orientation and conformation of the ligands at the binding site were determined by a recently-developed flexible ligand docking program (J. Wang, I.D. Kuntz and P.A. Kollman, submitted for publication). This docking program has been applied to twelve ligand-protein complexes with three to sixteen rotatable bonds and obtained the binding structures with root mean square deviations (rmsd) ranging from 0.64 Å to 2.01 Å from the corresponding crystal structures. For the biotin-avidin complex, the docked biotin structure has an rmsd of 0.68 Å from the respective crystal structure.

All molecular dynamics (MD) simulations in this work were carried out at 300 K and 1 atmosphere. The explicit solvent model TIP3P water[29] was used. The SHAKE procedure[30] was employed to constrain all solute bonds involving at least one hydrogen atom. The time step of the simulations was 1.5 fs with a dual cutoff of 10 Å and 17 Å for the nonbonded interactions. The nonbonded pairs were updated every 30 steps. All protein atoms within 18 Å of

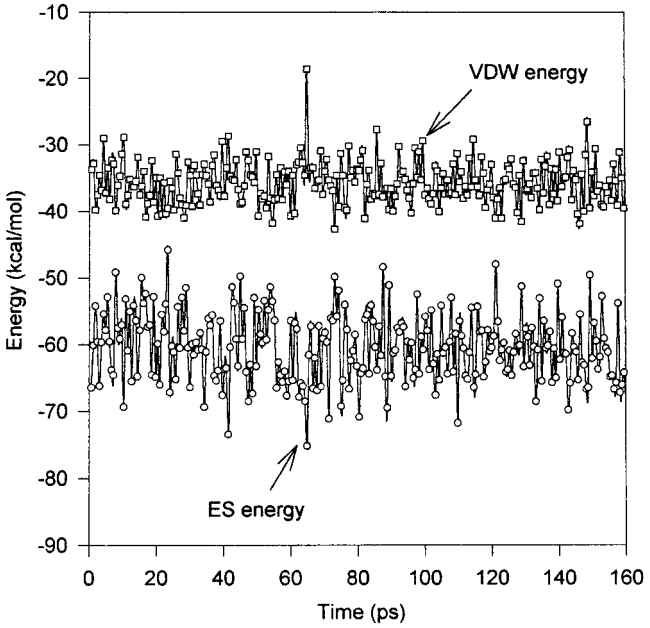

Fig. 1. Energy fluctuations over the molecular dynamics trajectory.

the center of mass of ligand were allowed to move during the MD runs. This was realized by the BELLY option in AMBER. In the BELLY (movable zone), a 20 kcal/mol/Å^2 position constraint was imposed for any residues between 15.5 Å and 18 Å from the center of mass of ligand. Prior to the MD runs, the system was minimized with harmonic position constraints for all protein heavy atoms in the BELLY. The constraints are 5,000, 1,000, 100 and 10 kcal/mol/Å^2. Subsequently, a cycle of minimization was done to relax all the atoms in the BELLY without constraints (except for the 20 kcal/Å^2 constraints for the residues between 15.5 Å and 18 Å from the center of mass of ligand). The maximum minimization steps were 10,000 and the convergence criterion for energy gradient was 0.5 kcal/mol/Å. In the MD simulation of an unbound ligand in aqueous solution, a position constraint (50 kcal/mol/Å^2) for the ligand heavy atom that is closest to the center of mass of the ligand was applied. During the date collection run, the interaction energies of the ligand with all solvent atoms (protein and water solvent) in the BELLY were calculated and printed out for use in the subsequent analysis.

One technical issue in applying the LIE approach is the length of the MD simulation. In principle, the longer the simulation, the more reliable the resultant binding free energy. In practice, a relatively short simulation seems adequately accurate in this case. Figure 1 illustrates fluctuation of the electrostatic energy and the vdW energy over the course of 160 ps MD simulation for the biotin-avidin complex. The electrostatic interaction energy has a fluctuation of about ±8 kcal/mol, while the vdW interaction energy fluctuates about ±5 kcal/mol. The mean energy difference between the first 80 ps and 160 ps is 0.32 kcal/mol for the electrostatic energy and 0.30 kcal/mol for the vdW energy. In general, we observe that vdW energy converges faster than electrostatic energy because of the short-range nature of the vdW interaction. In our several test cases, we observed that the mean energy is converged within 100 ps simulation. For example, in the biotin-avidin complex, the mean energy difference between the first 50 ps and the second 50 ps is 0.36 kcal/mol for the electrostatic energy and 0.31 kcal/mol for the vdW energy. Therefore, most MD simulations reported in this work started with a 50 ps equilibration and then were followed by a 100 ps production run.

Interaction Free Energy Calculations

In Eq. 1, $\alpha = 0.5$ is used for the electrostatic energy, while β is a parameter for the van der Waals interaction energy. We also explore an alternate way to estimate binding free energy based on the interaction energies. Since the interaction free energy can be estimated as

$$\begin{aligned}\Delta G_{\text{int}} &= -RT \ln \langle \exp(-E_{\text{int}}/RT) \rangle \\ &= -RT \ln \langle \exp[(-\overline{E}_{\text{int}} + \Sigma \Delta E^i_{\text{int}})/RT] \rangle \\ &= \overline{E}_{\text{int}} - RT \ln \langle \exp(-\Sigma \Delta E^i_{\text{int}}/RT) \rangle \\ &= \overline{E}_{\text{int}} - T\Delta S_{\text{int}} \end{aligned} \quad (3)$$

where the sum is over MD trajectory, $\overline{E}_{\text{int}} = < E_{\text{int}} > = \alpha < E^{ES}_{\text{int}} > + \beta < E^{vdW}_{\text{int}} >$ is an ensemble averaged interaction energy, $\Delta E^i_{\text{int}} = \alpha(E^{ES,i}_{\text{int}} - < E^{ES}_{\text{int}} >) + \beta (E^{vdW,i}_{\text{int}} - < E^{vdW}_{\text{int}} >)$, and ΔS_{int} is the entropic term related to fluctuation of the interaction energy, $\Delta S_{\text{int}} = \ln\langle \exp(-\Sigma \Delta E^i_{\text{int}}/RT) \rangle / R$, we can therefore estimate the binding free energy based on the interaction free energies of ligand-protein complex and the unbound ligand in solvent:

$$\Delta G_{bind} = \Delta G^{pro}_{\text{int}} - \Delta G^{sol}_{\text{int}} \quad (4)$$

Eq. 4 is the same as Eq. 1 except for the entropic term, ΔS_{int}. We found that this term can yield slightly better results in this set of ligands.

Free Energy Calculations

In this work, free energy calculation has been applied to some systems in order to compare with the LIE results. The free energy change between states A and B was estimated by means of thermodynamic integration method. By defining as a coupling parameter to link states 0 and 1, such that $H(\lambda = 0)$ and $H(\lambda = 1)$, where $H(\lambda = 0)$ and $H(\lambda = 1)$ are the Hamiltonian of states 0 and 1, respectively, the free energy difference between states 1 and 0 is

$$\Delta G = \Delta G_1 - \Delta G_0 = \int_0^1 \langle \partial V/\partial \lambda \rangle_\lambda \, d\lambda \quad (5)$$

where V is the energy function as defined in Eq. 2, and $\langle \rangle_\lambda$ is denoted as ensemble average at the intermediate state λ.

Using the thermodynamic cycle shown as following:

$$\begin{array}{ccc} P + L_1 & \xrightarrow{\Delta G_1} & P \cdot L_1 \\ \Delta G_{sol} \downarrow & & \downarrow \Delta G_{pro} \\ P + L_2 & \xrightarrow{\Delta G_2} & P \cdot L_2 \end{array}$$

The relative binding free energy of the protein receptor (P) with ligands L_1 and L_2 can be defined as:

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_{pro} - \Delta G_{sol} \quad (6)$$

where ΔG_1 and ΔG_2 are absolute free energies for binding the receptor by L_1 and L_2. ΔG_1 and ΔG_2 can be obtained from experimental measurements. ΔG_{sol} and ΔG_{pro} are free energy changes for ligand solvation and ligand-protein complexation, respectively. These two components can be calculated by means of thermodynamic integration method. One often uses the calculated $\Delta G_{pro} - \Delta G_{sol}$ values to compare with the experimentally observed $\Delta G_2 - \Delta G_1$ values to validate a theoretical model, or to predict relative binding free energies of a set of ligands against the same receptor if the experimental data are not available.

The TI simulations used the same protocol as the MD simulation except that only a 10 Å cutoff was used for the nonbonded interactions. Each thermodynamic integration (TI) simulation consisted of 51 windows, with 2,000 or 4,000 steps each for equilibration and data collection in each window. All TI calculations were performed with forward and backward runs. Fifty (50) ps equilibration was carried out prior to each direction of the TI runs.

Protocols of Protein Charged Residues

One important issue in calculating ligand-protein interaction energies is to determine the charge state of charged residues of the receptor. Since the length of MD or FEP simulation (~100 ps) is relatively short in this work and only a limited number of cap water molecules are used, adding counter-ions is not a good choice, because they cannot be adequately equilibrated. Ideally, one can use continuum solvation model such as Poison-Boltzmann method to estimate the charge states. Unfortunately, since the charge state is very sensitive to the choice of the dielectric constant of the medium, the continuum solvation model cannot always give a definite answer to the question of interest. Hence, no attempt has been made in this study to determine charge states of the charged residues through the continuum solvation model. Instead, we investigated two "simple" protocols. The first protocol turns off all protein charged residues beyond 14 Å of the center of mass of biotin, while keeping fully-charged residues within 14 Å. The second approach, on the other hand, turns off only the minimum number of outmost charged residues in order to keep net charge of zero for all the residues in the belly. Our test calculations for biotin, thiobiotin, and iminobiotin show that the binding free energy is not very sensitive to these two protocols. For instance, using $\alpha = 0.5$ and $\beta = 1$, the calculated binding free energy for biotin is −21.82 and −22.10 kcal/mol, respectively, with these two charge-residue protocols. Similarly, the binding free energy for iminobiotin is −14.25 and −13.84 kcal/mol with these two protocols, respectively. Hereafter we report our results based on the second protocol for charged residues. We should point out that the choice of charge state for protein charged residues is still an issue of debate and the approach employed here is only a temporary expedient.

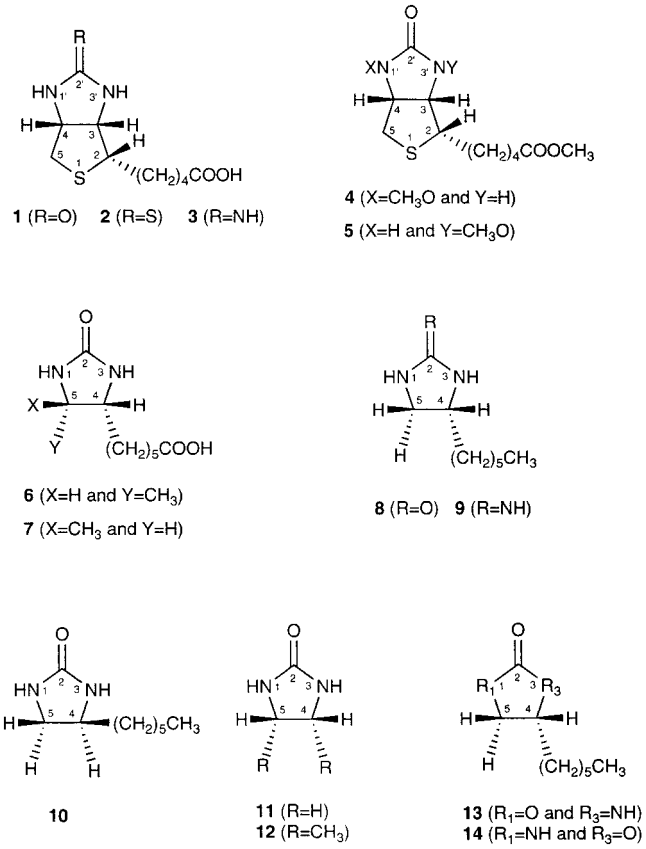

Fig. 2. Biotin analogs used in the binding free energy calculations.

Most simulations were carried out on HP 9000/735 workstations, while part of the TI simulations was carried out on HP-Convex Exemplar clusters.

Fourteen biotin analogues were selected in this study, as shown in Figure 2. These ligands have different sizes with either bicyclic five-member rings or only one five-member ring. Some ligands have two hydrogen bonds (H-bond) donors in the ureido ring, while others have one H-bond donor and one H-bond acceptor. The experimentally-observed binding free energies for these ligands range from −4.5 to −20.4 kcal/mol.[21]

RESULTS AND DISCUSSIONS

In this section, we discuss the binding free energy results calculated using either the interaction free energy or the LIE approaches and compare the calculated results

TABLE I. Rmsd From Avidin Crystal Structure for the Movable BELLY of Ligand-Avidin Complex at the MD Snapshots

	rmsd (Å)			
	1	4	10	14
After minimization	0.20	0.51	0.43	0.34
After equilibration	0.46	0.88	0.70	0.41
30 ps	0.54	1.24	0.64	0.82
60 ps	0.72	1.18	0.84	0.73
100 ps	0.62	1.21	0.77	0.79

with the corresponding experimental values. We also address some technical issues such as the effects of initial binding structure, ligand charge states, and protein backbone motions on the binding free energy calculations. Finally, an attempt has been made to understand why the interaction energy methods overestimate the binding free energies for ligand **4**, **5**, **13**, and **14**.

The Protein Structure During the Simulations

To warrant a detailed discussion of binding energetics, comparison of the protein structures in the simulations with the related experimental structure is crucial. A simple way to measure if the simulated structures drift from the experimental one is to calculate the root mean square deviation (rmsd) from the initial (crystal) structures for all heavy atoms in the moving BELLY region. Table I lists rmsd for several representative complex systems. Starting with the crystal structure of biotin-avidin complex, avidin stays close to its initial structure over the course of the simulation. The rmsd of the heavy atoms in the BELLY are 0.20 Å after minimization of the complex structure, 0.46 Å after 50 ps equilibration, 0.54 Å at the first 30 ps production run, and 0.62 at the 100 ps snapshot. Binding by **4** to avidin leads to a slightly large conformational change in avidin due to a bulky methoxyl substitution. The rmsd of the avidin heavy atoms in the **4**-avidin complex ranges from 0.51 to 1.24 Å. As seen in Table 1, for a small ligand, imidazolidine (**10**), the overall receptor structure also stays close to its starting structure. The rmsd of the avidin heavy atoms in the moving BELLY region are less than 0.84 Å. The similar rmsd is observed in the **14**-avidin complex as shown in Table I. Overall, the simulation protocol used in this work does not cause a significant drift of avidin away from the structure that was bound to biotin.

The Parameters for Determining the Binding Free Energies

Table II lists the calculated energy components, binding free energies as determined by the LIE method (Eq. 1) and the interaction free energy method (Eq. 4) together with the respective experimental data. Figure 3 shows the correlation of the calculated binding free energies with respect to the experimental values.

In most LIE calculations published in literature, $\alpha = 0.5$ has been used for the electrostatic interaction energy, while the value of β is an issue of debate. In this work, we determined $\beta = 1$ by comparing the interaction free energy calculation results on biotin (**1**), thiobiotin (**2**), and iminobiotin (**3**) with the respective experimental results. Since we started with the biotin-avidin crystal structure, the binding of these three ligands to avidin is expected to lead to the smallest change in the protein structures, consequently, the calculated energy components are less of an artifact. Hence, the determined β value from these sets of compounds is expected to be more reliable than that from other ligands. This is the main reason why we use these three ligands to determine the β value. We should point out that the binding free energy calculations at this level are aimed for the lead optimization, in which only few lead compounds are available. So selecting a small number of compounds as the calibration set is desired at this stage. We also note that the calculated binding energy results are quite sensitive to the β value used in the calculations. No extensive investigation has been carried out to find best β value for the avidin system.

Using $\alpha = 0.5$ and $\beta = 1.0$, both the LIE and interaction free energy methods yield the binding free energies in good correlation with the experimental results except for ligands **4**, **5**, **13**, and **14**. In general, the interaction free energy method (Eq. 4) appears to yield results slightly closer to the experimental binding free energy than the LIE approach due to the entropic term, $T<\Delta S_{int}>$. This term has contribtuions to binding of free energy in the range from $+0.05$ to -2.85 kcal/mol. As seen in Table 2, the calculated binding free energies using $\alpha = 0.5$ and $\beta = 0.167$ does not lead to as good agreement with experiments as the calculated values using $\alpha = 0.5$ and $\beta = 1.0$. Relative to the experimental binding free energies for ten out of the fourteen compounds (except **4**, **5**, **13**, and **14**), the standard deviations of the calculated binding free energies are 2.53 kcal/mol in the interaction free energy method with $\alpha = 0.5$ and $\beta = 1$, 3.56 kcal/mol in LIE using $\alpha = 0.5$ and $\beta = 1$, and 8.64 kcal/mol in LIE using $\alpha = 0.5$ and $\beta = 0.167$. Obviously, a better correlation between the calculated and experimental binding free energy results can be obtained by adjusting β, but no attempts have been made to do so in this work.

In our test runs, we also carried out calculations for benzamidin-trypsin complex, the system Åqvist has studied using the GROMOS force field.[14] Using the same protocol as for the biotin-analogs, the energy components calculated using the AMBER force fields were listed in Table III together with Åqvist's results. In general, the energy components calculated from the AMBER and GROMOS force fields are different. For example, the electrostatic interaction energy for the benzamidine-trypsin complex is -114.14 kcal/mol in AMBER and -125.38 in GROMOS, the vdW interaction energy for benzamidine in water is -9.82 kcal/mol in AMBER and -7.96 in GROMOS. The differences of these energy components, on the other hand, are similar to each other despite that different force fields were used in the calculations. The net electrostatic interaction energy is -12.22 kcal/mol in

TABLE II. Binding Free Energies for Neutral Ligand Binding to Avidin With $\alpha = 0.5$ and $\beta = 1$ (kcal/mol)

Ligand	$\langle V_{el} \rangle$	$\langle V_{VW} \rangle$	$-\langle T\Delta S \rangle$	ΔG_{cald}	ΔE^{b}_{LIA}	ΔE^{c}_{LIA}	ΔG_{expt}
1 (biotin)							
ΔE[a]	−4.68	−18.04	1.07	−21.82	−23.72	−7.69	−20.4
Unbound	−26.34 ± 0.08	−19.53 ± 0.06	−6.41 ± 0.23	−52.29 ± 0.18			
Bound	−31.20 ± 0.08	−37.57 ± 0.30	−5.34 ± 0.90	−74.11 ± 0.55			
2 (thiobiotin)							
ΔE[a]	−3.93	−14.95	0.89	−17.99	−18.86	−6.42	−16.9
Unbound	−22.36 ± 0.08	−21.14 ± 0.06	−5.31 ± 0.08	−48.81 ± 0.06			
Bound	−26.29 ± 0.14	−36.09 ± 0.03	−4.42 ± 0.25	−66.80 ± 0.11			
3 (iminobiotin)							
ΔE[a]	2.28	−18.31	1.69	−14.25	−16.03	−0.77	−14.3
Unbound	−28.46 ± 0.14	−21.17 ± 0.07	−9.07 ± 0.22	−58.70 ± 0.16			
Bound	−26.18 ± 0.51	−39.48 ± 0.07	−7.38 ± 0.12	−73.05 ± 0.38			
4 (1′-N-methoxycarbonylbiotin methyl ester)							
ΔE[a]	0.30	−20.82	0.65	−19.86	−20.52	−3.18	−8.8
Unbound	−21.25 ± 0.17	−26.66 ± 0.04	−5.72 ± 0.03	−53.63 ± 0.15			
Bound	−20.95 ± 0.13	−47.48 ± 0.04	−5.07 ± 0.09	−73.49 ± 0.06			
5 (3′-N-methoxycarbonylbiotin methyl ester)							
ΔE[a]	−3.46	−15.90	0.73	−18.63	−19.36	−6.11	−12.2
Unbound	−21.14 ± 0.03	−26.48 ± 0.05	−6.45 ± 0.29	−54.07 ± 0.27			
Bound	−24.60 ± 0.04	−42.38 ± 0.26	−5.72 ± 0.11	−72.70 ± 0.15			
6 (1-desthiobiotin)							
ΔE[a]	−2.93	−12.65	−0.05	−15.62	−15.48	−5.04	−14.0
Unbound	−13.78 ± 0.04	−22.62 ± 0.06	−4.26 ± 0.09	−40.67 ± 0.06			
Bound	−16.71 ± 0.09	−35.27 ± 0.35	−4.31 ± 0.11	−56.29 ± 0.33			
7 (d-desthiobiotin)							
ΔE[a]	−2.93	−15.97	2.85	−16.05	−18.90	−5.60	−16.5
Unbound	−13.56 ± 0.03	−22.84 ± 0.03	−6.06 ± 0.11	−42.46 ± 0.10			
Bound	−16.49 ± 0.05	−38.81 ± 0.10	−3.21 ± 0.16	−58.51 ± 0.05			
8 (D-4-n-hexylimidazolidone)							
ΔE[a]	−3.45	−11.73	0.72	−14.26	−15.18	−5.41	−11.1
Unbound	−13.39 ± 0.10	−18.89 ± 0.03	−4.21 ± 0.07	−36.49 ± 0.08			
Bound	−16.84 ± 0.05	−30.62 ± 0.05	−3.29 ± 0.09	−50.75 ± 0.10			
9 (D-4-n-hexyl-2-iminoimidazolidine)							
ΔE[a]	−0.94	−12.61	0.67	−12.89	−13.55	−1.21	−7.4
Unbound	−15.44 ± 0.12	−17.49 ± 0.07	−4.37 ± 0.13	−37.30 ± 0.07			
Bound	−16.38 ± 0.02	−30.10 ± 0.16	−3.70 ± 0.10	−50.19 ± 0.19			
10 (L-4-n-hexylimidazolidone)							
ΔE[a]	−3.30	−8.44	2.43	−9.31	−11.74	−4.71	−9.1
Unbound	−13.48 ± 0.06	−19.12 ± 0.03	−5.64 ± 0.09	−38.24 ± 0.06			
Bound	−16.78 ± 0.03	−27.56 ± 0.15	−3.21 ± 0.10	−47.55 ± 0.05			
11 (Imidazolidone)							
ΔE[a]	−4.06	−5.42	1.77	−7.71	−9.48	−4.97	−4.5
Unbound	−12.72 ± 0.08	−8.06 ± 0.02	−3.45 ± 0.05	−24.23 ± 0.03			
Bound	−16.78 ± 0.03	−13.48 ± 0.01	−1.68 ± 0.03	−31.94 ± 0.02			
12 (dl-4,5-dimethylimidazolidone)							
ΔE[a]	−3.26	−7.03	1.11	−9.18	−10.29	−4.43	−6.4
Unbound	−13.25 ± 0.05	−11.45 ± 0.02	−3.78 ± 0.04	−28.48 ± 0.04			
Bound	−16.51 ± 0.02	−18.48 ± 0.02	−2.67 ± 0.14	−37.66 ± 0.10			
13 (D-4-n-hexyloxazolidone)							
ΔE[a]	1.75	−14.83	0.73	−12.35	−13.08	−4.23	−5.0
Unbound	−12.76 ± 0.05	−19.57 ± 0.05	−4.23 ± 0.07	−36.55 ± 0.08			
Bound	−11.01 ± 0.05	−34.40 ± 0.16	−3.50 ± 0.03	−48.90 ± 0.14			
14 (D-5-n-hexyloxazolidone)							
ΔE[a]	−1.27	−13.70	0.78	−14.20	−14.97	−3.56	−7.4
Unbound	−12.83 ± 0.05	−18.84 ± 0.03	−4.18 ± 0.20	−35.85 ± 0.20			
Bound	−14.10 ± 0.04	−32.54 ± 0.30	−3.40 ± 0.08	−50.05 ± 0.19			

[a] $\Delta E = \langle E_{bound} \rangle - \langle E_{unbound} \rangle$.
[b] $\beta = 1.0$.
[c] $\beta = 0.167$.

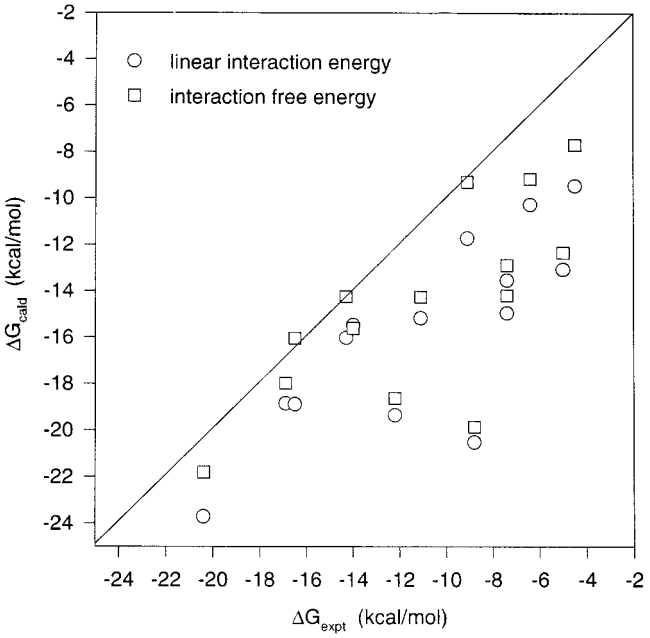

Fig. 3. The correlation between the calculated binding free energies and the corresponding experimental results.

TABLE III. The Calculated Energy Components (kcal/mol) for Benzamidine With Trypsin

	$\langle V_{el} \rangle$	$\langle V_{vdW} \rangle$	ΔG^{a}_{cald}	ΔG_{expt}
AMBER				
unbound	−101.92 ± 0.50	−9.82 ± 0.18		
bound	−114.14 ± 0.10	−20.35 ± 0.09		
ΔE	−12.22	−10.53	−7.87	−6.4
GROMOS[14]				
unbound	−115.18 ± 0.12	−7.96 ± 0.06		
bound	−125.38 ± 0.02	−20.87 ± 0.13		
ΔE	−10.20	−12.91	−7.26	−6.4

[a]Using $\alpha = 0.5$ and $\beta = 0.167$ and without electrostatic energy correction.[14]

AMBER and −10.20 kcal/mol in GROMOS, whereas the net vdW interaction energy is −10.53 kcal/mol in AMBER and −12.91 in GROMOS. Using $\alpha = 0.5$ and $\beta = 0.167$, the calculated binding free energy for benzamidine with trypsin is −7.87 kcal/mol in AMBER and −7.26 kcal/mol in GROMOS, which compares favorably with the experimental binding value of −6.4 kcal/mol.

The similar binding free energies from both force fields indicate that the need to use $\beta = 1$ for avidin-biotin rather than Åqvist's $\beta = 0.167$ is likely due to the differences in protein rather than the differences in the force fields used in the calculations. A further discussion of the dependence of β on the protein systems will be presented elsewhere.[31]

Charged Ligands vs Neutral Ligands

Calculating the binding free energies of charged ligands is a challenge. This is because charged ligands correspond to a large desolvation energy and ligand-protein-complex energy. Calculation of absolute binding free energies thus inherits a large numerical error due to the fact that binding free energy is the difference between these two large numbers. This problem becomes more obvious when one attempts to estimate binding free energies of a set of compounds mixed with neutral and charged states, as the intrinsic errors for these two kinds of ligand are different. In this section, we discuss the effect of ligand charge state on binding free energy.

Table IV compares the calculated binding free energies for biotin (**1**), thiobiotin (**2**), and iminobiotin (**3**) with either protonated or deprotonated states. Using $\alpha = 0.5$ and $\beta = 1$, the calculated binding free energies for the neutral ligands (protonated state) are in good agreement with the corresponding experimental results. Although the calculated binding free energies for charged ligands **1**, **2**, and **3** (deprotonated states) are in the correct order, their relative energies deviate from the corresponding experimental results more than those of the neutral compounds. For example, using the interaction free energy, the relative binding free energy between biotin and thiobiotin is –3.7 kcal/mol in experiment and −3.8 kcal/mol for neutral ligands, but it is only −1.4 kcal/mol for charged ligands. Such error (−3.7 − (−1.4) = 2.4 kcal/mol) makes it more difficult to make reliable predictions. Therefore, in spite of the fact that using the anionic states rather than the protonated states seems appropriate for **1**, **2**, and **3** because the carboxylic acid should be ionized at the pH (∼5) in the experimental measurements, the possible artifacts in the calculations suggest that it appears reasonable to use the protonated states for **1**, **2**, **3** and the other-COOH analogues in the LIE-like approaches. Therefore, we used the results for neutral ligands **1**, **2**, **3**, **6**, and **7** in Table 2 for comparison with the others neutral ligands. We emphasize that there will be situations where one must use the appropriately charged ligand for reliable results (e.g. benzamidine-trypsin).

Dependence of the Calculated Binding Free Energy on Starting Structures

In ligand binding free energy calculations, the docking structure is the starting point for minimization and subsequent molecular dynamics simulation. Because the interaction energy approach uses ensemble-averaged energies to estimate binding free energies, an interesting question to ask is if the calculated interaction energies are sensitive to the docking structures. Here we chose the **13**-avidin complex, whose binding energy was overestimated by more than 7 kcal/mol in the calculations, as an example to address this issue.

Starting with the lowest-energy docking structure of **13**-avidin, we manually generated several structures by either rotating the ureido ring of **13** by about ±40° or translating its center of mass by 2.0 and 3.5 Å. The generated binding orientations were visually inspected so that there is no vdW clash between the ligand and the residues in the binding pocket. Choosing the docking

TABLE IV. Binding Free Energies of Charged and Protonated Ligands in Avidin With $\alpha = 0.5$ and $\beta = 1$ (kcal/mol)

Ligands	$\langle V_{el} \rangle$	$\langle V_{VW} \rangle$	$\langle T\Delta S \rangle$	ΔG_{cald}	ΔE_{LIA}	ΔG_{expt}
Charged ligands						
(1) Biotin						
ΔE[a]	−0.36	−20.64	1.34	−19.65	−20.28	−20.6
Unbound	−92.04 ± 0.05	−12.75 ± 0.06	−16.24 ± 0.09	−121.04 ± 0.08		
Bound	−92.40 ± 0.13	−33.39 ± 0.09	−14.90 ± 0.07	−141.69 ± 0.12		
(2) Thiobiotin						
ΔE[a]	1.39	−18.67	1.76	−18.30	−20.16	−16.9
Unbound	−87.54 ± 0.05	−14.42 ± 0.03	−13.97 ± 0.22	−115.93 ± 0.24		
Bound	−88.93 ± 0.19	−33.09 ± 0.15	−12.21 ± 0.12	−134.23 ± 0.08		
(3) Iminobiotin						
ΔE[a]	5.59	−20.33	2.88	−11.86	−14.74	−14.3
Unbound	−97.09 ± 0.23	−11.89 ± 0.02	−16.95 ± 0.26	−125.93 ± 0.18		
Bound	−91.50 ± 0.46	−32.22 ± 0.07	−14.07 ± 0.13	−137.79 ± 0.32		
Neutral ligands (protonated)						
(1) Biotin						
ΔE[a]	−4.68	−18.04	1.07	−21.82	−23.72	−20.4
Unbound	−26.34 ± 0.08	−19.53 ± 0.06	−6.41 ± 0.23	−52.29 ± 0.18		
Bound	−31.20 ± 0.08	−37.57 ± 0.30	−5.34 ± 0.90	−74.11 ± 0.55		
(2) Thiobiotin						
ΔE[a]	−3.93	−14.95	0.89	−17.99	−18.86	−16.9
Unbound	−22.36 ± 0.08	−21.14 ± 0.06	−5.31 ± 0.08	−48.81 ± 0.06		
Bound	−26.29 ± 0.14	−36.09 ± 0.03	−4.42 ± 0.25	−66.80 ± 0.11		
(3) Iminobiotin						
ΔE[a]	2.28	−18.31	1.69	−14.25	−16.03	−14.3
Unbound	−28.46 ± 0.14	−21.17 ± 0.07	−9.07 ± 0.22	−58.70 ± 0.16		
Bound	−26.18 ± 0.51	−39.48 ± 0.07	−7.38 ± 0.12	−73.05 ± 0.38		

[a] $\Delta E = \langle E_{bound} \rangle - \langle E_{unbound} \rangle$.

TABLE V. The Structures and Binding Free Energies With Respect to Different Starting Orientation of 13 in Avidin Binding Pocket

Number	rmsd (Å) initial	rmsd (Å) after min.	rmsd (Å) after eq.	$\Delta G_{binding}$ (kcal/mol)
a	0.00	0.00	0.00	−12.35
b	2.50	1.98	0.50	−12.77
c	3.67	3.65	0.43	
d	4.30	3.82	0.61	−12.61
e	4.56	4.31	0.71	−12.42

structure as reference, the root mean square deviations of these structures range from 2.2 Å to 4.5 Å and have relatively higher energies than the docking structure (results not shown in detail). After a series of minimizations and 50 ps MD equilibration, most of these structures were not much different from the one that was started from the docked structure. The rmsd of these structures from the docked structure range from 1.5 Å to 3.9 Å after minimization, while the rmsd after 50 ps MD equilibration is less than 1 Å, as illustrated in Table V. The difference among the calculated binding free energies for three of these structures is less than 0.5 kcal/mol. This indicates that the orientation energy landscape of **13** in avidin binding pocket is relatively flat and molecular dynamics simulation samples the orientation space reasonably well. Therefore, the binding free energy calculation using MD-based method is not very sensitive to the starting orientation of a ligand in this case.

The Effect of Protein Flexibility on Binding Free Energy

In our MD simulations so far, a harmonic position constraint of 20 kcal/mol/Å^2 was imposed only on the residues in the BELLY that are 15–18 Å away from the center mass of a ligand. An 18 Å cap of water molecules was used to mimic the solvent effect near the binding pocket. The limited number of waters might not be enough to keep the protein in "proper" motion, and furthermore, it may distort the protein motion or ligand-protein structure in the movable BELLY. In order to investigate the effect of protein flexibility on binding free energy calculation, we carried out MD simulations by imposing a harmonic position constraint for the avidin backbone heavy atoms in the BELLY. The force constant used was 100 kcal/mol/Å^2.

The resultant binding free energies are listed in Table VI for selected ligands. As seen in Table VI, the limitation of protein motion does not affect the binding free energy on biotin (**1**), because the calculation used the biotin-avidin crystal structure in which ligand is in perfect match with the binding site. In the **4**-avidin complex, the harmonic position constraint on backbone heavy atoms in the movable BELLY improves the binding free energy results. The binding free energy for **4** is –8.8 kcal/mol in experiment, –19.86 in the interaction free energy calculation without

TABLE VI. Binding Free Energy of Selected Neutral Ligands in Avidin Calculated With a Harmonic Position Constraint (100 kcal/mol/Å^2) for All Avidin Backbone Heavy Atoms in the BELLY With $\alpha = 0.5$ and $\beta = 1$ (kcal/mol)

Ligand	$\langle V_{el} \rangle$	$\langle V_{VW} \rangle$	$\langle T\Delta S \rangle$	ΔG^a_{rigid}	ΔG^b_{flex}	ΔG_{expt}
1 (biotin)						
ΔE	−5.19	−17.72	1.38	−21.52	−21.82	−20.4
4 (1′-N-methoxycarbonylbiotin methyl ester)						
ΔE	7.64	−16.34	1.62	−7.07	−19.86	−8.8
5 (3′-N-methoxycarbonylbiotin methyl ester)						
ΔE	0.41	−16.47	−0.77	−16.83	−18.63	−12.2
6 (l-desthiobiotin)						
ΔE	−5.09	−13.39	1.04	−17.43	−15.62	−14.0
9 (D-4-n-hexylimidazolidone)						
ΔE	−5.35	−11.32	0.97	−15.70	−12.89	−11.1
11 (imidazolidone)						
ΔE	−5.45	−4.49	1.76	−8.18	−7.71	−4.5
12 (dl-4,5-dimethylimidazolidone)						
ΔE	−4.41	−5.56	1.66	−8.31	−9.81	−6.4
13 (D-4-n-hexyloxazolidone)						
ΔE	4.65	−9.92	1.12	−4.16	−12.35	−5.0
14 (D-5-n-hexyloxazolidone)						
ΔE	−3.99	−12.89	0.96	−15.91	−14.20	−7.4

[a]The binding free energies calculated using the backbone sontraints.
[b]The binding energies calculated without backbone constraints.

the constraint of the belly residues and −7.07 kcal/mol with the constraint. The constraint also leads to better binding free energy results in the **13**-avidin complex. The binding free energy for **13**-avidin changed from −12.35 kcal/mol in the calculations without the constraint to −4.16 kcal/mol with the constraint, which is in good agreement with the experimental value of −5.0 kcal/mol.

Unfortunately, limitation of the backbone motion does not lead to a consistent improvement in binding free energies for all ligands studied in this work. In **8** and **14**, for example, the calculated relative binding free energies were worse than the results calculated without backbone constraint. The relative binding free energy between **4** and **5** are −3.4 kcal/mol in experiment, +1.2 kcal/mol without the constraint and −9.76 kcal/mol with the constraint. Similarly, the relative binding free energy between **13** and **14** is –2.4 kcal/mol in experiment, –1.85 kcal/mol for the calculation without the constraint, while it is –11.75 kcal/mol for the calculation with the constraint. The same observation holds for ligand **11** and **12**.

Overall, calculations using rigid backbone for avidin residues do not give rise to a consistent improvement in the binding free energies. This study also indicates that the binding free energy results are sensitive to the protocol of protein flexibility used in calculation. A careful examination of this effect is necessary in a calculation.

FEP vs LIE: A Case Study

Since limitation of protein motion in the calculations does not consistently improve the binding free energy results for **4**, **5**, **13,** and **14**, here, we carried out free energy perturbation calculations for **4** and **5** in order to understand these calculated results. Using thermodynamic integration method, we examine the relative binding free energies between **1** and **4** and between **1** and **5**. The calculated free energy changes are listed in Table VII. The free energy calculations for **1** → **4** and for **1** → **5** in vacuum and solution were converged within 600 ps of simulation. The hystereses between forward and backward runs are less than 0.5 kcal/mol. The simulations for the ligand-avidin complexes, however, are more difficult to converge. Even with 600 ps of simulation, the hysteresis is 1.67 kcal/mol for **1** → **4** and 2.53 kcal/mol for **1** → **5**. Since the relative binding free energies of interest are more than 8 kcal/mol experimentally, which is larger than the uncertainty in the calculation, the following discussion is thus based on the 600 ps simulation results.

As seen in Table VII, the relative binding free energies calculated from TI are in much better agreement with the respective experimental results than the interaction energy methods. For instance, the experimentally-observed relative binding free energy for **1** → **4** is 11.6 kcal/mol, the TI simulation yields 10.31 kcal/mol, the interaction free energy and the LIE methods predict the relative binding free energy to be 1.96 and 3.20 kcal/mol, respectively. Similarly, for **1** → **5**, the relative binding free energy is 8.2 kcal/mol in experiment, 7.03 kcal/mol in the TI calculation, 3.19 kcal/mol in the interaction free energy calculation, and 4.36 kcal/mol in the LIE calculation. Clearly, the TI calculation has better performance than the interaction energy approaches in terms of predicting relative ligand binding free energies.

Why does the TI calculation yield better results than the interaction energy approaches? First of all, looking at the biotin-avidin (**1**-avidin) complex structure in detail as shown in the stereoview of the 100 ps snapshot (Fig. 4), the

TABLE VII. Calculated Free Energies (kcal/mol) for Perturbing 1 → 4 and 1 → 5 in Vacuum, in Aqueous Solution and in Avidin

	free energy (ΔG)[a]						
Length (ps)	Total	EL	14EL	vdW	14vdW	badh	pmf
1 → 4							
In vacuum							
600	82.44 ± 0.17	−72.83	147.10	7.00	7.43	−3.40	−2.90
In solution							
300	96.54 ± 0.56						
600	94.99 ± 0.14	−61.23	144.19	10.09	7.45	−0.34	−5.17
In protein							
300	105.66 ± 2.75						
600	105.30 ± 1.66	−56.73	142.40	14.08	6.72	4.96	−6.13
ΔΔG(sol/vac)	12.46 ± 0.22						
ΔΔG(pro/vac)	22.86 ± 1.67						
ΔΔG(pro/sol)	10.31 ± 1.67	4.50	−1.79	3.99	−0.73	5.34	−0.96
$\Delta\Delta G_{expt}$	11.60						
1 → 5							
In vacuum							
600	106.11 ± 0.43	−41.58	131.96	8.96	7.70	2.09	−3.02
In solution							
300	115.80 ± 1.86						
600	115.09 ± 0.24	−34.80	133.08	15.14	7.13	−0.25	−5.44
In protein							
300	121.81 ± 5.65						
600	122.12 ± 2.53	−32.90	132.02	18.87	7.55	2.01	−5.43
ΔΔG(sol/vac)	8.98 ± 0.49						
ΔΔG(pro/vac)	16.01 ± 2.57						
ΔΔG(pro/sol)	7.03 ± 2.54	1.90	−1.06	3.73	0.42	2.26	−0.01
$\Delta\Delta G_{expt}$	8.2						

[a]EL: electrostatic energy contribution to the free energy; 14EL: 1–4 electrostatic energy contribution to the free energy; vdW: van der Waals energy contribution to the free energy; 14vdW: 1–4 van der Waals energy contribution to the free energy; badh: internal energy including bond stretching, bond angle bending and torsional angle energies of the ligand contribution to the free energy; and pmf: free energy correction for changing bond length of the perturbed group.

ureido oxygen of protonated biotin (**1**) forms a hydrogen bond network with Asn12, Ser16, and Tyr33. Two NH groups in the ureido ring function as H-bond donors to atom OD1 of Asn118 and atom OG1 of Thr35. All these residues are deeply buried inside the binding pocket and their solvent accessibility is null in the presence of biotin. The tetrahydrothiophenic ring and aliphatic chain are surrounded by aromatic residues Phe79, Trp97, and by Trp110 from the other subunit. The terminal carboxylic acid also has H-bonds with Ala38, Thr40, and Ser73. The electrostatic interaction energy of −31.20 kcal/mol in biotin-avidin complex seems to be dominated by these H-bonds. These H-bonds, on the other hand, exist for an unbound biotin in solvent water as well. The net gain of electrostatic energy in binding is thus only about −4 kcal/mol, whereas the van der Waals energy is −18.04 kcal/mol, as seen in Table II. This is consistent with the free energy perturbation results carried out by Miyamoto and Kollman.

Relative to biotin, **4** and **5** are from the substitution of the hydrogen with a methoxyl group at either 1′ or 3′ position of biotin (Fig. 2). The terminal acidic group in biotin is also changed to ester. The structural deviation of avidin heavy atoms within the BELLY is 1.2 and 1.0 Å, respectively, upon binding of **4** and **5** to avidin. In the **4**-avidin complex (Fig. 5), two H-bonds between ureido oxygen and Ser16 and between ureido oxygen and Tyr33 remain as that in the biotin-avidin complex, but the H-bond between the ureido oxygen and Asn12 is broken due to the bulky substitution. The H-bond between HN in the ureido ring and Thr35 is also broken, whereas Tyr33 forms an H-bond as an H-bond donor with nitrogen at 1′ position. In the terminal region, the H-bond between the hydrogen in carboxylic acid of biotin and Thr39 is broken due to the methyl substitution. The ester group exposes to solvent and has no close contact (more than 3 Å) with protein residues. Therefore, relative to the biotin-avidin complex structure, avidin has a noticeable conformation change in the binding site upon substituting the N-H with a bulky methoxyl group. In the other words, the residues in the binding site must reorganize themselves in order to create a cavity for **4**. This part of the free energy penalty is only related to protein itself, which is not explicitly included in the interaction energy calculations for **4**-avidin. In the free energy calculation, not only the interaction energy, but also the protein reorganization energy penalty

Fig. 4. Stereoview of **1**-avidin (biotin-avidin) complex.

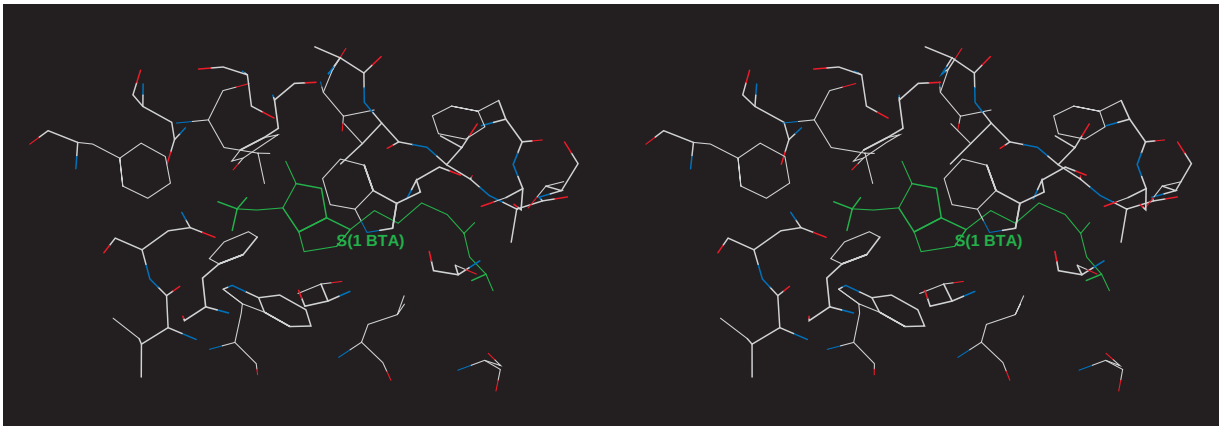

Fig. 5. Stereoview of **4**-avidin (1′-N-methoxycarbonylbiotinmethylester-avidin) complex.

and other energies are included, so TI provides a more complete description on the ligand-protein interaction than the interaction energy approaches. Accordingly, TI can produce better results in binding free energy calculation than the interaction energy methods.

Note that the binding free energy of **4**-avidin is correctly predicted in the LIE calculations using position constraint for backbone atoms of avidin. One may argue that the results of the TI simulation are fortuitous, because the TI simulations for **1** → **4** had limited sampling, in which protein binding site is "effectively" rigid just like in the MD simulation with the constraint. This argument is not correct. First of all, TI predicted the correct relative binding free energy for **1** → **5** as well, while the interaction energy calculations with the backbone constraint failed to do so. Secondly, the energy components obtained from these two calculations are different. For example, in the TI calculation for **1** → **4**, the relative binding free energy of 11.60 kcal/mol consists of 4.50 kcal/mol in electrostatic energy, 3.99 kcal/mol in vdW energy and 1.86 kcal/mol in 1–4 electrostatic energy, 1–4 vdW energy and internal energy altogether. In the interaction free energy calculation with the backbone constraint, on the other hand, 12.83 kcal/mol in the electrostatic energy and 1.38 kcal/mol in the vdW energy mainly contribute to the relative binding energy of 14.35 kcal/mol. The different energy component contributions to the binding free energy by these two methods clearly show that the TI calculations are different from the interaction free energy method, though their relative binding free energy results are similar.

The similar discussion might be applied to the binding of **13** and **14** to avidin, in which LIE overestimates the absolute binding free energy for **13** and **14** by more than 7 kcal/mol relative to the respective experimental values. Also, the conformational entropy of the floppy -$(CH_2)_5$-CH_3 group in **13** and **14** (also in **8** and **9**) might contribute to the overestimation in the binding free energies for these molecules, because inclusion of this would stabilize the unbound state and reduce the magnitude of the calculated binding free energies.

It seems that the interaction energy approach cannot handle well those ligands that will induce noticeable conformational change of protein during the binding pro-

cess. When a bulky substitution with hetero-atoms occurs inside the binding pocket, the protein residues need to reorganize their conformation in order to create a cavity for the ligand binding. In order to describe such ligand-protein interaction, free energy calculation such as thermodynamic integration or free energy perturbation methods is still appropriate.

CONCLUSIONS

The binding of biotin analogous to avidin has been studied by means of molecular dynamics simulations. Two molecular dynamics-based methods to estimate binding free energy are examined in this work. The first method, often called the linear interaction energy (LIE) method, employs the interaction energies between a ligand and its surrounding in either bound or unbound states to calculate the binding free energy. The second method utilizes the interaction free energy to estimate the ligand binding free energy. The interaction free energy is calculated based on the interaction energy in the first approach. For 10 out of 14 biotin analogues, both the LIE and the interaction free energy methods lead to binding free energies that correlate with the corresponding experimental values. The interaction energy approach appears to give slightly better results than LIE due to the inclusion of the entropy contribution in the binding free energy calculations. For the remaining four ligands, both methods are found to overestimate the binding free energies by more than 7 kcal/mol. The origin of the overestimation is attributed to the conformational change of the binding site induced by binding. A substitution with heteroatoms occurs inside the binding site and forces the residues in the pocket to reorganize in order to create a cavity for the ligand binding. The free energy price for the protein reorganization is not explicitly included in the LIE or the interaction free energy calculations. Thermodynamic integration method that includes this part of energy can reproduce the experimentally observed relative binding free energies. The TI calculations support that the internal energy change of protein is important for the binding by these ligands to avidin. Nonetheless, the interaction energy method has been proven to be useful to quickly estimate ligand binding free energy.

The van der Waals coefficient, β, in Eq. 1 and Eq. 3 is very important in estimating binding free energies for ligand-protein complexes. Based on our own and Åqvist and his coworkers' similar results for benzamidine-trypsin, it appears that the variation of β is not simply due to different force fields used in the calculations. Also, the fact that Paulsen and Ornstein's P450cam results found a similar β ($\sim$1.0) to that found here and very different from Åqvist et al's β ($\sim$0.167) makes it appear that the optimal β is protein-dependent. A subsequent paper will address the dependence of protein-ligand interactions.[31]

ACKNOWLEDGEMENTS

This research was supported in part by the grants GM-29072 and GM-56609 (P.I., E. Arnold) to P.A.K. Part of this investigation was completed using the facilities in the UCSF Computer Graphics Lab, T. Ferrin, director, supported by RR-1081 from the NIH. Gratitude is expressed to Prof. Irwing D. Kuntz for stimulating discussions. Thanks to Dr. Kechuan Tu for arranging computation hours on a HP 9000/735 workstation. J.W. thanks Drs. L. Wang, Carlos Simmerling, and Y. Duan and Mr. Wei Wang for helpful discussions. J.W. gratefully acknowledges the Natural Science and Engineering Council of Canada for a postdoctoral fellowship to carry out part of this research.

REFERENCES

1. Kuntz ID, Meng EC, Shoichet BK. Structure-based molecular design. Accounts of Chem Res 1994;27:117–123.
2. Rosenfeld R, Vajda S, Delisi C. Flexible docking and design. Annu Rev Biophys Biomol Struct 1995;24:677–700.
3. Böhm H-J. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des 1992;6:61–78.
4. Williams DH. et al. Toward the semiquantitative estimation of binding constants. J Am Chem Soc 1991;113:7020–7030.
5. Head RD. et al. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. J Am Chem Soc 1996;118:3959–3969.
6. Eldridge MD, Murray CW, Auton TR, Paolini GV. Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 1997;11:425–445.
7. Ewing T, Kuntz ID. Critical evaluation of search algorithms for automated molecular docking and database screening. J Comp Chem 1997;18:1175–1189.
8. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
9. 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.
10. Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: parallel application of AutoDock 2.4. J Comput Aided Mol Des 1996;10:293–304.
11. Kollman PA. Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 1993;93:2395–2417.
12. Beveridge DL, Dicapua FM. Free energy via molecular simulations: Application to chemical and biochemical system. Annu Rev Biophys Biophys Chem 1989;18:431–492.
13. van Gunsteren WF. Methods for calculation of free energies and binding constants: Success and problems. In: Gunsteren WFv, Weiner PK, editors. Computer simulation of biomolecular systems: theoretical and experimental applications. Netherlands: Escom Science Publishers; 1989. p. 27–59.
14. Åqvist J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. J Comp Chem 1996;17:1587–1597.
15. Åqvist J, Hansson T. On the validity of electrostatic linear response in polar solvents. J Phys Chem 1996;100:9512–9521.
16. Hansson T, Åqvist J. Estimation of binding free energies for hiv proteinase inhibitors by molecular dynamics simulations. Protein Eng. 1995;8:1137–1144.
17. Åqvist J, Medina C, Samuelsson JE. New method for predicting binding affinity in computer-aided drug design. Protein Eng 1994;7:385–391.
18. Paulsen MD, Ornstein RL. Binding free energy calculations for P450cam-substrate complexes. Protein Eng 1996;9:567–571.
19. Jones-Hertzog DK, Jorgensen WL. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. J Med Chem 1997;40:1539–1549.
20. Heney G, Orr GA. The purification of avidin and its derivatives on 2-iminbiotin-6-aminohexylsepharose 4B. Anal Biochem 1981;114:92–96.
21. Green NM. Avidin. Adv Protein Chem 1975;29:85–133.

22. Green NM. Avidin and Streptavidin. London: Academic Press; 1990. p. 51–67.
23. Wilchek M, Bayer EA. Avidin biotin technology 10 years on - Has it lived up to its expectations. Trends Biochem Sci 1989;14:408–412.
24. Livnah O, Bayer EA, Wilchek M, Sussman JL. 3-Dimensional structures of avidin and the avidin biotin complex. Proc Nat Acad Sci USA 1993;90:5076–5080.
25. Pugliese L, Coda A, Malcovati M, Bolognesi M. 3-Dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2.7 angstrom resolution. J Mol Biol 1993;231:698–710.
26. Miyamoto S, Kollman PA. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. Proteins 1993;16:226–245.
27. Miyamoto S, Kollman PA. What determines the strength of noncovalent association of ligands to proteins in aqueous solution? Proc Natl Acad Sci USA 1993;90:8402–8406.
28. Pearlman DA. *et al.* AMBER4.1. San Francisco: University of California;1995.
29. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. TIP3P water model. J Chem Phys 1983;79:926–935.
30. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equation of motion of a system with constraint: molecular dynamics of n-alkanes. J Comp Phys 1977;23:327–341.
31. Wang W, Wang J, Kollman PA. What determines the van der Waals coefficient β in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations. Proteins, in press.