New Method for Calculating the Absolute Free Energy of Binding: The Effect of a Mobile Loop on the Avidin/Biotin Complex

Ignacio J. General, Ralitsa Dragomirova, and Hagai Meirovitch*

Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, 3059 BST3, Pittsburgh, Pennsylvania 15260, United States

Received: August 13, 2010; Revised Manuscript Received: November 23, 2010

Hypothetical scanning molecular dynamics (HSMD) is a relatively new method for calculating the absolute free energy and entropy. HSMD is extended here for the first time for calculating the absolute free energy of binding, ΔA^0, as applied to the avidin−biotin complex. With HSMD the ligand is built (more accurately reconstructed) from nothing in solvent and in the protein, in contrast to the commonly used methods where the ligand is annihilated (by thermodynamic integration) in these environments. Therefore, the end-point problem encountered with the latter methods does not exist with HSMD and the need for restraints is avoided. Also, the entropy of the ligand and water in both environments is obtained directly as a byproduct of the simulation. The binding mechanism of biotin to avidin involves a mobile loop that is expected to be in an open conformation in unbound avidin, which is changed to a closed one upon binding, that is, the loop moves to cover biotin in the active site. The contribution of the loop's conformational change to the total free energy of binding is calculated here for the first time. Our result, $\Delta A^0 = -24.9 \pm 7$ covers the experimental value −20.7 kcal/mol within the error bars.

I. Introduction

The noncovalent association of molecules is central to many biological processes, such as the action of hormones, the recognition of antigens by the immune system, the catalysis of chemical reactions by enzymes, and the action of drugs. Therefore, development of simulation methods for calculating the affinity (free energy) of molecular binding is important for both academic and practical reasons as they will help in elucidating the mechanisms of complex biological processes, which might be used in rational drug design and might lead to other therapeutic means.

In addition to the available simplified and fast scoring functions that enable one to screen large databases of ligands in drug design, it is imperative to also devise highly accurate methods for calculating the free energy of binding based on detailed molecular interactions and rigorous statistical mechanics; such methods are required in the refinement stage of the screening procedure. Of special interest in this category are methods for calculating the *absolute* (standard) free energy of binding of a (small) ligand to a protein or a DNA.

Indeed, in the last 30 years, a great deal of work has been done in this direction, where various techniques have been developed and applied to a wide range of problems (refs 1−11 and references cited therein). A central approach is based on thermodynamic cycles, where the interactions between the ligand and its environment are decreased to zero in both, the active site and the bulk solution (water), using thermodynamic integration (TI) or free energy perturbation (FEP) procedures.

Implementing this approach, which is sometimes called the double annihilation method (DAM)[1−5] is not straightforward. During the final stages of TI, the ligand leaves the active site and starts wandering within the volume (incluing the protein), which makes it extremely difficult to obtain converged results; also in most of the DAM applications the effect of the standard state has been ignored (ref 12 as recent example). Still, this method has been used successfully in recent years,[8,13] where in a later study[13] the effect of the standard state has been considered. The end-point problem has been rigorously solved by adding restraints that hold the ligand in the active site; the corresponding bias introduced is later removed by releasing the restraints. Because of the additional integration (e.g., TI) step involved, this procedure is sometimes called the double decoupling method (DDM).[3−5] However, application of DDM can become an elaborate process due to the (typical) use of several harmonic restraints and the need to optimize their force constants.[14] This approach has been developed systematically in the last 15 years,[9,7,15] where various implementation issues have been improved and a large number of complexes have been successfully studied (review by Deng and Roux[9]).

In spite of this success, it should be noted that every technique has limitations, and those of DAM have been specified above. Likewise, one can envisage potential limitations of DDM,[16] for example, in treating large flexible ligands;[5] also, there is always the risk that applying external forces will impose irreversible changes on the complex (e.g., disruption of hydrogen bonds), which can affect the free energy significantly.[17]

Therefore, it is desirable to develop *alternative* techniques with unique properties for calculating the free energy of binding. A potential such method is the hypothetical scanning molecular dynamics (HSMD) (or HSMC when Monte Carlo replaces MD)[18−22] − a general exact technique for calculating the absolute entropy and free energy, which has been developed in our group in the last several years. HSMD is extended here for the first time to handle protein−ligand complexes; we hope that it will enrich the field by increasing the arsenal of tools for calculating binding affinities.

As will be explained later, with HSMD, which is an *exact* method, the ligand in both environments is built (more exactly reconstructed) from nothing rather than annihilated or decoupled

* To whom correspondence should be addressed. E-mail: hagaim@pitt.edu; Phone: 412-648-3338.

10.1021/jp1076752
Published on Web 12/15/2010

(as with DAM and DDM); therefore the end-point problem encountered with DAM does not exist and the need for applying restraints is avoided. Also, the individual contributions of the ligand and water to the entropy are obtained in both environments as byproducts of the simulation and the method is particularly suitable to handle large ligands; calculation of the ligand entropies is important in rational drug design. Thus, HSMD constitutes a new method independent of existing techniques for calculating the absolute free energy of binding.

Our method is applied here to the avidin−biotin complex, which has been studied extensively both by experimental and computational techniques. The entire methodology consists of three main components, (1) application of HSMD, (2) thermodynamic integration (TI), and (3) calculation of the effect of a mobile loop which covers the bound biotin (this effect has been ignored in most of previous computational work). To introduce all these components and check their numerical performance, we apply HSMD-TI in this work to a somewhat simplified model of avidin, where its coordinates are held fixed in their X-ray structure positions (the template). The present results will shed light on the quality of this model, and will constitute reference values for comparison with future HSMD-TI results for the avidin−biotin complex based on more detailed models; preliminary results for a moving template are presented as well.

I.1. Avidin−Biotin Complex. HSMD is applied here to the avidin−biotin complex. Hen egg-white avidin is a tetrameric glycoprotein composed of 4 × 128 residues, which can bind up to four biotin molecules. The experimental free energy of binding is one of the highest in nature, $\Delta A^0 = -20.7$ kcal/mol ($K_d = 10^{-15}$ M),[23] a property that has been exploited to devise powerful tools for affinity chromatography, biochemical assays, and many other applications.[24−26] Biotin binds to the end of the β-barrel formed by each subunit of avidin and as is shown by the crystal structures of the complex, the bound biotin is covered by the 12-residue surface loop, 35−46 (connecting strands 3 and 4, thus denoted L3,4);[27,28] this closed conformation is expected to contribute significantly to the binding affinity. In fact, in streptavidin − a protein with a similar structure and binding affinity ($\Delta A^0 = \sim -18$ kcal/mol) deletion of the corresponding 8-residue loop (45−52) via circular mutation decreases the binding affinity by approximately 10 kcal/mol.[29] Also, in computer experiments this (covering) loop disturbed the pulling of biotin from avidin and had to be removed.[30] Because the conformational changes of this loop are expected to affect the free energy, we discuss them in some detail below.

The conformation of L3,4 in apo avidin is less clear. In the early X-ray structures of Pugliese at al.[27,31] (based on crystals belonging to tetragonal space group $P4_22_12$) the apo (protein data bank (pdb) 1avd) and holo (1ave) structures were found to be similar (rmsd of 0.57 Å) meaning that the loops share approximately the same (closed) conformation. However, Pugliese et al. comment that a closed-loop conformation leads to a narrow aperture that would prevent biotin from entering the active site in apo avidin. This situation, which is not expected to exist in solution, is attributed to the effect of crystal contacts. Indeed, in a later study by Livnah et al.[28] (using crystals of the space group, $P2_12_12$) the apo and holo structures (with 2.6 and 3 Å resolution, respectively) were found to be the same besides L3,4 which appears disordered in apo avidin (2avi). A disordered L3,4 in apo avidin was also observed by Nardeone et al.[32] [$P2_12_12$ and 2.2 Å resolution (1rav)]. Finally, in a recent high-resolution study of apo avidin by Repo et al.[33] [$P2_12_12$; 1.48 Å resolution (1vyo)] both the closed and open conformations of L3,4 are observed in the electron density maps, where the closed-loop conformation is similar to those found previously.

The open conformation of L3,4 in 1vyo diverges from the closed-loop conformation at residue I34 and reunites with the closed-loop conformation at K45. However, only the two starting residues, T35−A36 and the three ending ones, N42−K45, could be built into the electron density map of the open-loop conformation (the conformation of V37−S41 of the open loop could not be traced). These studies suggest that the conformation of L3,4 in apo avidin depends on the crystal and it is plausible to assume that in solution L3,4 is open an appreciable part of the time to enable binding of ligands. (Correspondingly, in most of the crystal structures of apo streptavidin, L3,4 exists in the open structure.[34]) On the basis of this assumption we shall consider the effect of the open and closed structures in the calculation of the free energy of binding.

II. Theory and Methodology

II.1. Modeling the Avidin−Biotin System. In this initial study of the free energy of binding we are mainly interested in checking the performance of HSMD, therefore (as mentioned earlier) the modeling of avidin/biotin/water is somewhat limited. As described later, application of HSMD requires MD simulations of biotin in solvent (water) and in the active site of avidin. For water we apply the TIP3P model[35] and for biotin we use the force-field parameters derived by Israilev et al.[30] where the total charge is zero, whereas for the experimental measurements at pH∼5 the carboxylic acid should be ionized.[23] However, a charged biotin would require considering long-range electrostatic effects which we seek to avoid in this initial study.

Conformational search study of biotin in vacuum and water (section III.1) has yielded structures similar to that of bound biotin (in 1avd), which thus was chosen as the initial unbound structure for MD simulations; it was solvated by a sphere of 794 water molecules with bulk water density (0.0350 $Å^{-3}$). These waters are kept within the sphere (of 18 Å radius) by a flat-welled harmonic potential with a spring constant of 10 kcal mol^{-1} $Å^{-2}$ (i.e., if the distance of a water molecule from the center is greater than 18 Å, the potential is turned on; otherwise, it is off); this system is called the solvent system.

On the protein side, we started with the pdb structure 1avd,[27] which was relaxed by energy minimization, where the heavy atoms were restrained to their crystal structure positions by harmonic forces with a constant of 5 kcal/(mol Å). However, treating this tetrameric biotin−avidin structure would be computationally expensive; therefore, we consider only a spherical part of the molecule consisting of the residues within a distance of 18 Å from the center of the binding site of subunit B. This reduces the number of atoms (including hydrogens) from ∼8000 to 2082, where the optimized positions of the latter atoms define a fixed template (of +2 charge), which is not moved by MD. Next, we defined a water sphere of 18 Å radius centered in the binding site, which was filled with 355 TIP3P waters with the normal density of bulk water. As in the solvent system, water molecules are kept within the sphere by a harmonic potential of 10 kcal mol^{-1} $Å^{-2}$. This system will be called the protein system.

MD simulations of the protein and solvent systems were performed with the *TINKER 5.0* package[36] using the AMBER 99 force field,[37] where Lys, Arg, Glu, Asp, and His are charged; the temperature was kept at 300 K with the Berendsen thermostat, based on a time constant of 1.0 ps.[38] The RATTLE algorithm kept bonds of hydrogens constant, which allowed using a 2 fs time step, and no boundary conditions or cut-offs

on interactions were applied. Notice that only the biotin and water molecules move in the simulation, whereas the template−biotin and template−water interactions (with the fixed template) are taken into account. Thus both the protein and solvent systems are described by the canonical (*NVT*) ensemble.

As pointed out above, we also seek to take into account the open conformation of L3,4 in apo avidin. However, unlike streptavidin, only the coordinates of the end residues 35−36 and 42−45 are available. Therefore, we carried out a conformational search (without biotin) consisting of five MD runs where the end residues of the initial loop conformations have the corresponding pdb coordinates (1vyo). The three final conformations with the coordinates of the end residues closest to the pdb values were chosen to represent the open conformation of L3,4.

II.2. Absolute Free Energy of Binding. Because our systems are defined in the *NVT* ensemble, we rely on the derivation of ref 5 where the absolute Helmholtz free energy of binding ΔA^0 is

$$\Delta A^0 = -k_B T \ln \frac{\bar{Z}_{PL,N} Z_{0,N}}{8\pi^2 V^0 \bar{Z}_{P,N} \bar{Z}_{L,N}} \quad (1)$$

P, L, and N stand for protein, ligand, and the number of solvent molecules (water), respectively. Thus, $\bar{Z}_{PL,N}$, $\bar{Z}_{P,N}$, and $\bar{Z}_{L,N}$ are the conformational partition functions of the complex, protein, and ligand all in water; $Z_{0,N}$ is the partition function of N water molecules in the volume. The bar means that P and L in $\bar{Z}$ are defined by *internal coordinates*; $8\pi^2$ comes from integrations over the three (angular) external coordinates, where $V^0 = 1660$ Å^3 is the standard volume. Expressing ΔA^0 in terms of $\bar{Z}$ is convenient as with HSMD only the internal entropies of the ligand are calculated. Still, notice that the ligand moves in the active site and we shall see later that in the framework of HSMD this effect can be treated *exactly* by calculating the probability of the six external coordinates of the ligand (in the active site but not in solvent). However, because the present model the volume of the active site remains constant, this effect is small and is thus calculated approximately by evaluating the product $\Omega_a V_a$, which is added to eq 2 below; V_a is the volume available for biotin's reference atom in the active site and Ω_a is an average angular contribution due to integration of the three Euler angles.[4] This is the only approximation applied in our methodology. ΔA^0 is expressed in terms of configurational (Helmholtz) free energies, F, and an additional term

$$\begin{aligned} \Delta A^0 = (\bar{F}_{PL,N} - \bar{F}_{P,N}) - (\bar{F}_{L,N} - F_{0,N}) + \\ k_B T \ln (8\pi^2 V^0/\Omega_a V_a) = \\ \Delta F_p - \Delta F_{sol} + k_B T \ln (8\pi^2 V^0/\Omega_a V_a) \end{aligned} \quad (2)$$

ΔF_p and ΔF_{sol} are free energy differences defined for the protein environment and the ligand in solvent, respectively, which are calculated by HSMD. Also, the absolute Gibbs free energy, $\Delta G^0 \sim \Delta A^0$ because $\Delta G^0 = \Delta A^0 + P^0 \Delta \bar{V}_{PL}$ where the (ignored) $P^0 \Delta \bar{V}_{PL}$ is expected to be small.[4,7]

II.3. HSMC(D) Method. The HSMC(D) method enables one to calculate the absolute entropy and free energy from a sample generated by MD,[39,40] Monte Carlo[41] (MC), or any other simulation technique. HSMC(D) is based on the fact that a system configuration can, in principle, be generated *exactly* also by a step-by-step (growth) procedure, where particles are added gradually to an initially empty volume using transition probabilities (TPs). A trivial example is an *ideal chain* of N bonds on a square lattice, that is a chain without the excluded volume interaction. An ideal chain can be simulated *exactly* by the dynamical MC method where the entropy, however, is unknown; alternatively, a chain configuration, i, can be constructed from nothing as a random walk where a bond's direction (out of four) is chosen with TP $= {}^1/_4$ at each step; here the Boltzmann probability is known, $P_i^B = ({}^1/_4)^N$ and thus the entropy is known as well. Clearly large samples constructed by MC or as random walks are equivalent, in the sense that they lead to the same thermodynamic averages and fluctuations.

However, for a complex system, application of this exact growth procedure will in general be very inefficient. Still, relying on its equivalence with MC (MD), one can assume that a given MC (MD) sample has rather been generated by the exact growth method, which enables one to reconstruct each conformation i by calculating the TP densities that *hypothetically* were used to create it step-by-step (from now on we shall omit in most cases the letters MC).

In practice, the product of the TPs leads to an approximation, P_i for the correct Boltzmann probability P_i^B, where from P_i various free energy functionals (F) can be defined. Whereas the TPs of HSMD are stochastic in nature (calculated by MD simulations), all of the system interactions are taken into account; in this respect HSMD can be viewed as exact,[18] where the only approximation involved is due to insufficient MD sampling for calculating the TPs. HSMD has unique features: it provides rigorous lower and upper bounds for F, which enable one to determine the accuracy from HSMD results alone without the need to know the correct answer. Furthermore, F can be obtained from a very small sample and in principle even from *any* single conformation (e.g., see results for argon in ref 18). The HSMD methodology has been developed systematically, as applied to systems of increasing complexity. The initial (HSMC) calculations of liquid argon, TIP3P water,[18] self-avoiding walks, and polyglycine molecules[22] were found to be highly accurate by comparisons with results obtained by other well established techniques; these results also demonstrate the above theoretical predictions related to F.[22] Below we describe how HSMD is applied for calculating ΔF_{sol} of eq 2.

One starts by generating an MD sample of size n_s of biotin soaked in N water molecules. The reconstruction of biotin is done in internal coordinates which are the $K = 28$ dihedral and bond angles ordered along the chain where their set is denoted $[\alpha_k]$, $k = 1...K$ [the bond lengths are ignored because they do not contribute to entropy differences.[19−21] The Jacobians related to the bond angles were found in previous work and in the present one to get canceled in entropy differences; therefore, for simplicity they are ignored in the following discussion]. In practice, one denotes the ordered heavy atoms (and one hydrogen) along the chain by $k'=1...{}^K/_2 = 14$, that is $k = 2k'$; the dihedral and bond angles, $k - 1$ and k define the position of atom k'.

First one calculates the variability range $\Delta\alpha_k = \alpha_k(\max) - \alpha_k(\min)$, where $\alpha_k(\max)$ and $\alpha_k(\min)$ are the maximum and minimum values of α_k found in the sample. Each of the configurations (frames) $([\alpha_k],\mathbf{x}^N)$ (denoted i for brevity) of the sample is reconstructed in two stages, where the biotin structure is reconstructed first followed by the reconstruction of the water configuration $\mathbf{x}^N$. Because the position of atom k' is defined by a dihedral and a bond angle one has to calculate their TP

simultaneously. Thus, at step k' ($k = 2k'$) of stage 1, the $k-2$ angles $\alpha_{k-2}\cdots\alpha_1$ have already been reconstructed and the TP density of $\alpha_{k-1}\alpha_k$, $\rho(\alpha_{k-1}\alpha_k|\alpha_{k-2},\cdots,\alpha_1)$, is calculated from an MD run, where the *entire future* of biotin and water is moved (i.e., biotin's atoms $k',k'+1,\ldots^{K}/_{2}$ and their connected hydrogens, and the water coordinates $\mathbf{x}^N$), whereas the *past* (biotin's atoms 1,2, ..., $k'-1$ and their connected hydrogens) are held fixed at their values in conformation i. By considering a future conformation every 6 fs, a sample of size n_f is generated. Two small segments (bins) $\delta\alpha_{k-1}$ and $\delta\alpha_k$ are centered at $\alpha_{k-1}(i)$ and $\alpha_k(i)$ respectively and the number of *simultaneous* visits, n_{visit}, of the future chain to these two bins during the simulation is calculated; one obtains[19−22]

$$\rho_{biotin}(\alpha_{k-1}\alpha_k|\alpha_{k-2},\cdots,\alpha_1) \approx \rho^{HS}(\alpha_{k-1}\alpha_k|\alpha_{k-2},\cdots,\alpha_1) = n_{visit}/[n_f\delta\alpha_{k-1}\delta\alpha_k] \quad (3)$$

where $\rho^{HS}(\alpha_{k-1}\alpha_k|\alpha_{k-2},\cdots,\alpha_1)$ becomes exact for very large n_f ($n_f \rightarrow \infty$) and very small bins ($\delta\alpha_{k-1}$, $\delta\alpha_k \rightarrow 0$). This means that in practice $\rho^{HS}(\alpha_{k-1}\alpha_k|\alpha_{k-2},\cdots,\alpha_1)$ will be somewhat approximate due to insufficient future sampling (finite n_f) and relatively large bin sizes. The corresponding probability density related to the biotin's conformation is

$$\rho^{HS}(\alpha_K,\cdots,\alpha_1) = \rho^{HS}([\alpha_k]) = \prod_{k=2,2}^{K}\rho^{HS}(\alpha_{k-1}\alpha_k|\alpha_{k-2},\cdots,\alpha_1) \quad (4)$$

$\rho^{HS}([\alpha_k])$ defines an approximate entropy functional, denoted S^{A}_{btn}, which can be shown (using Jensen's inequality) to constitute a *rigorous* upper bound for the correct, S_{btn},[18]

$$S^{A}_{btn} = -k_B\int_m \rho^{B}_{btn}([\alpha_k])\ln\rho^{HS}([\alpha_k])d[\alpha_K] \quad (5)$$

$\rho^{B}_{btn}([\alpha_k])$ is the Boltzmann probability density of $[\alpha_k]$. ($S^{A}_{btn} \geq S_{btn}$ is also known as the Gibbs' inequality). Being an upper bound suggests that S^{A}_{btn} will decrease as the approximation improves. S^{A}_{btn} is estimated by $\bar{S}^{A}_{btn}$ from an MD (Boltzmann) sample of size n_s.

$$\bar{S}^{A}_{btn} = -(k_B/n_s)\sum_{t=1}^{n_s}\ln\rho^{HS}(t) \quad (6)$$

The entropy of biotin in the active site is calculated in the same way, where the reconstruction by MD simulations also depends on the water−template (denoted tmpl) and biotin−template interactions. We denote the entropies of biotin in the protein and solvent (water) environments by $S^{A}_{btn}(p)$ and $S^{A}_{btn}(sol)$ respectively, where our main interest is in their *converged* difference, ΔS_{btn}, which is expected to be the *exact* difference within the statistical error,

$$\Delta S_{btn} = S^{A}_{btn}(sol) - S^{A}_{btn}(p),\ \text{converged} \quad (7)$$

Thus, we calculate $S^{A}_{btn}(sol)$ and $S^{A}_{btn}(p)$ for increasing values of n_f and decreasing bins, verifying that both entropies decrease monotonically as the approximation improves, that is, both approach to the correct values from above. Typically, the convergence of ΔS^{A}_{btn} is much faster than that of the individual entropies, because of cancellation of comparable systematic errors in $S^{A}_{btn}(sol)$ and $S^{A}_{btn}(p)$. Thus, one can obtain ΔS_{btn} in any desired accuracy within a given statistical error where the required convergence is reached by increasing n_f and decreasing the bin sizes (the statistical error also depends on the sample size, n_s and other simulation parameters); the range of errors obtained in our previous work is 0.2−1 kcal/mol. Therefore, HSMD is considered to be an exact procedure. As pointed out earlier the probability of the reference atom of the ligand in the active site should be calculated; this is discussed later in section III.3.

II.4. HSMD-TI. In principle, one can reconstruct also the water configuration.[18] However, this procedure is time consuming and would be necessary for calculating the absolute values of the four components, $\bar{F}_{PL,N}$, $\bar{F}_{P,N}$, and so forth in eq 2. On the other hand, the water contribution to the *differences*, ΔF_p and ΔF_{sol}, can be calculated for each of the n_s configurations by a thermodynamic integration (TI) procedure in which the already reconstructed biotin conformation $[\alpha_k]$ is kept fixed and the water−biotin interactions are increased gradually from zero to their full values. In practice, however, it is easier to decrease these interactions to zero, where the charges are eliminated first (using a parameter $\lambda \rightarrow 0$) followed by the elimination of the Lennard−Jones (LJ) potential, which leads (in solvent) to the contributions, $F^{TI}_{water}([\alpha_k],sol,ch)$ and $F^{TI}_{water}([\alpha_k],sol,LJ)$.[20,21] Averaging over the n_s configurations yields $F^{TI}_{water}(sol,ch)$, $F^{TI}_{water}(sol,LJ)$ and their sum, $F^{TI}_{water}(sol)$. Corresponding functionals (which are the contribution of water to ΔF_p, eq 1) are defined in the protein environment [$F^{TI}_{water}(p,ch)$,$F^{TI}_{water}(p,LJ)$, etc.]. Again, only water−biotin interactions are eliminated (whereas water−template, biotin−template, and water−water interactions remain intact and the L3,4 loop is closed). These functionals are denoted by p instead of sol.

One can also calculate the difference in the water energy between the end points (denoted by $\lambda = 1$ and 0) where in solvent $E_{water}(sol) = E_{water-water}(sol,1) + E_{water-btn}(sol,1) - E_{water-water}(sol,0)$, which leads to the change in the water entropy $TS_{water}(sol) = E_{water}(sol) - F^{TI}_{water}(sol)$. In the same way, one defines $E_{water}(p)$ and $TS_{water}(p)$, where $E_{water-tmpl}(p,1)$ and $E_{water-tmpl}(p,0)$ should also be included. This calculation is possible because biotin is kept fixed during TI.

We also seek to calculate the difference in free energy between the closed and open conformations of the L3,4 loop. Because the loop constitutes a part of the fixed template, and template−template interactions are ignored, only the change in loop−water interactions is considered in the calculation of ΔF_{loop}. Thus, ΔF_{loop} is calculated by the same TI procedure described above applied to each fixed loop structure (where the rest of the template without biotin is the same). It is convenient to identify the closed loop with the protein and the open loop with solvent. Thus, as above, one obtains $F^{TI}_{loop}(p,ch)$, $F^{TI}_{loop}(p,LJ)$ and their sum, $F^{TI}_{loop}(p)$, and the corresponding $F^{TI}_{loop}(sol,ch)$, $F^{TI}_{loop}(sol,LJ)$, and $F^{TI}_{loop}(sol)$. Notice that the latter three functionals are averages over three open conformations obtained by the conformational search mentioned earlier. Thus, $\Delta F_{loop} = F^{TI}_{loop}(sol) - F^{TI}_{loop}(p)$. In view of all of these TI calculations, our method is referred to as HSMD-TI. Adding the free energy of the loop to eq 2, leads to

$$\Delta A^{o} = k_B T \ln (8\pi^2 V^0/\Omega_a V_a) + F^{TI}_{loop}(p,closed) + E_{btn-btn}(p) + E_{btn-tmpl}(p) - TS_{btn}(p) + F^{TI}_{water}(p) - [F^{TI}_{loop}(sol,open) + E_{btn-btn}(sol) - TS_{btn}(sol) + F^{TI}_{water}(sol)] = k_B T \ln (8\pi^2 V^0/\Omega_a V_a) + [F^{TI}_{loop}(p,closed) + F_{btn}(p) + F^{TI}_{water}(p)] - [F^{TI}_{loop}(sol,open) + F_{btn}(sol) + F^{TI}_{water}(sol)] \quad (8)$$

where $E_{btn-btn}(p)$, $E_{btn-btn}(sol)$, and $E_{btn-tmpl}(p)$ are the averages (over n_s configurations) of the intrabiotin and biotin−template energies; these energies together with S_{btn} define F_{btn}. The two branches in brackets in eq 8 define a thermodynamic cycle (depicted in Figure 1).

III. Results and Discussion

III.1. Preliminary Studies. Before simulating biotin in water, two questions should be answered, (1) is the water sphere of radius 18 Å large enough, and (2) whether the most stable biotin structures in solution differ from the bound structure. To answer (1) positively, we increased the radius up to 40 Å and simulated the corresponding systems by MD to find that changes in the average biotin−water interaction energy are less than 1%. To answer (2), we carried out an extensive conformational search of biotin in vacuum using the MC minimization method,[42] which is expected to reach the global minimum for this relatively small molecule. Altogether 40 runs (each of 10^4 minimizations) were performed starting from 20 different initial structures and using constant and varying temperatures. The lowest energies ($E_{btn-btn}$) obtained range from −102 and to −119 kcal/mol as compared to $E_{btn-btn} = -84$ kcal/mol of the active site structure. Each of the ten lowest energy structures was then solvated with water and was subjected to a 200 ps MD simulation at 300 K. However, all of the final conformations were found to be similar to the active site structure, which therefore was used as the initial structure in the production MD runs in solvent.

The production (MD) run in solvent was started with energy minimization of the biotin−water system using the L-BFGS algorithm[43] until a gradient of 0.01 was reached; then the system was heated gradually to 300 K, and an MD equilibration run of 1 ns was performed followed by 2 ns production run from which a set of $n_s = 88$ equally spaced frames were taken for later reconstruction and TI calculations by HSMD-TI.

To test the validity of the protein model described earlier [denoted (1)] we defined two additional larger models where in model (2) the whole tetramer is considered (with a 18 Å water sphere), and model (3) consisting of the tetramer with 40 Å water sphere; for these models we calculated the various energy components. The differences are very small, for example $E_{btn-tmpl} = 77.2$, 78.4, and 78.4 kcal/mol and $E_{btn-water} = 10.5$, 10.5, and 10.4 kcal/mol for systems (1), (2), and (3) respectively. We also decreased the charge of the template Histidines to zero to find again that the effect is small, for example $E_{btn-tmpl}$ was changed from 77.2 to76.9 kcal/mol. The 2 ns production run of biotin and water in the protein (with a fixed template) was carried out in the same way described above for biotin in solvent leading again to $n_s = 88$ equally spaced frames for later reconstruction and TI runs.

III.2. HSMD Results. We reconstructed $n_s = 88$ biotin conformations (denoted i) in both, the solvent and the protein, which has led to results of $\Delta\alpha_k$, $k = 1...K = 28$. Each reconstruction step (out of 14) starts from conformation i with a 10 ps equilibration and a 60 ps production run where a future biotin conformation is stored every 6 fs for a later analysis; thus, the total sample for each step consists of $n_f = 10^4$ future conformations. The number of counts, n_{visit} (eq 3) for each pair of bins is calculated leading to $TP_{k'}$, where the product of the 14 TPs is the distribution, ρ^{HS} (eq 4), which leads to the entropy, S^A (eqs 5 and 6). Results for S^A are presented in Table 1 as a function of n_f and bin size $\delta = \Delta\alpha_k/l$ where $l = 40$, 80, and 100; as expected, these results decrease systematically as the approximation improves, that is, with increasing n_f (and for the protein also with decreasing δ) and they seem to be close to convergence for $n_f = 6 \times 10^3 - 10^4$. The corresponding results

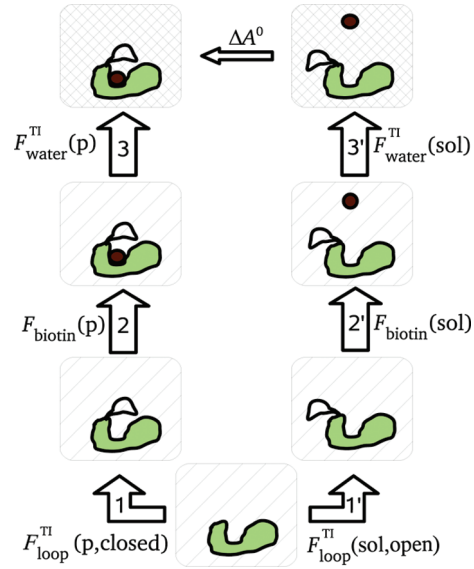

Figure 1. Thermodynamic cycle defined by eq **8**. 1 and 1′, change in the loop structures; 2 and 2′, reconstruction of biotin; 3 and 3′, TI of biotin−water interactions.

TABLE 1: HSMD Results (in kcal/mol) for the Entropy of Biotin, TS^A_{btn} (eq 6) and for Entropy Differences $T\Delta S^A_{btn}$ (eq 7) between Biotin in Solvent (sol) and Bound to Avidin (p) at $T = 300$[a]

bin size, δ	n_f	TS^A_{btn}(sol)	TS^A_{btn}(p)	$T\Delta S^A_{btn}$
$\Delta\alpha_k/40$	1000	54.1	46.4	7.7
	2000	50.6	43.3	7.3
	4000	49.3	42.2	7.1
	6000	48.9	41.8	7.1
	8000	48.6	41.7	6.9
	10 000	**48.5**	**41.6**	**6.9**
$\Delta\alpha_k/80$	1000	54.1	46.3	7.8
	2000	50.6	43.0	7.6
	4000	49.3	42.0	7.3
	6000	48.8	41.6	7.2
	8000	48.6	41.5	7.1
	10 000	**48.5**	**41.4**	**7.1**
$\Delta\alpha_k/100$	1000	54.1	46.3	7.8
	2000	50.6	43.0	7.6
	4000	49.3	41.9	7.4
	6000	48.8	41.5	7.3
	8000	48.6	41.3	7.2
	10 000	**48.5**	**41.3**	**4.0**
converged				**7.1 ± 0.6**

[a] The results were obtained by reconstructing 88 biotin structures selected homogeneously from MD samples of 2 ns for biotin in solvent and in the protein. The results are calculated as functions of $\delta = \Delta\alpha_k/l$ and n_f (eq 3) − the bin and sample size of the future chains, respectively. S^A_{btn} is defined up to an additive constant that is expected to be the same for both environments. The maximal statistical error for TS^A_{btn} is 0.7 kcal/mol. The results for $n_f = 10^4$ are bold-faced.

TABLE 2: Results in kcal/mol for the Energy, Free Energy, and Entropy of Water[a]

	F^{TI}_{water}(ch)	F^{TI}_{water}(LJ)	F^{TI}_{water}	E_{water}	TS_{water}
protein	-10.3 ± 0.2	3.6 ± 0.1	-6.7 ± 0.2	1.1 ± 3	7.8 ± 3
solvent	-23.6 ± 0.4	4.5 ± 0.1	-19.1 ± 0.4	-43.5 ± 6	-24.4 ± 6

[a] F^{TI}_{water}(ch) and F^{TI}_{water}(LJ) are free energies calculated by TI where the biotin−water electrostatic and Lennard−Jones interactions, respectively are increased from $\lambda = 0$ to $\lambda = 1$; F^{TI}_{water} is their sum. $E_{water} = E_{water}(\lambda = 1) - E_{water}(\lambda = 0)$ and TS_{water} is the corresponding change in entropy (for details see text).

for ΔS^A already converge to 7.1 ± 0.6 kcal/mol, which is thus considered to be the exact result. The significantly smaller entropy in the protein is expected due to the constraints imposed by the fixed template (stages 2 and 2′ in Figure 1).

III.3. TI Results. For each of the 88 frames of the two samples, we carried out a TI procedure where biotin−water interactions are turned off gradually for a *fixed* biotin structure (and a fixed template). This annihilation was performed using a parameter λ, where the electrostatic interactions were removed first followed by the removal of the LJ interactions (in the presence of zero electrostatic interactions). This TI is based on soft-core potentials defined by parameters λ and δ, where $\delta = 3$ Å^2.[44] For each type of interaction 22 values of λ (windows) were used ($\lambda = 0.95, 0.90, 0.85, ..., 0.10, 0.05, 0.03, 0.01, 0.00$). As in our previous study,[20,21] integration step i starts by minimizing the last structure of the previous step, $i - 1$, according to the potential energy of the current $\lambda(t)$, followed by a 5 ps equilibration, which uses the set of velocities of the last $i - 1$ structure (for $\lambda = 0.95$ we used a longer equilibration of 15 ps); after equilibration, a 20 ps production run is performed.

The results for the various functionals related to F^{TI}_{water} are summarized in Table 2 (also stages 3 and 3′ in Figure 1), which demonstrates that the errors [sd/($88^{1/2}$), where sd is standard deviation] are relatively small. This is important because the present system is the largest studied thus far by HSMD-TI. In Table 2, we also present results for E_{water} and TS_{water}. In solvent the water−biotin interaction ($\lambda = 1$) is relatively strong decreasing the energy of water by 43 kcal/mol (from $\lambda = 0$) and as one would expect, the entropy of water is also decreased significantly by 24 kcal/mol, that is, TS_{water} is ∼55% of E_{water}. In the protein, biotin is secluded to a large extent from water which is characterized by small positive E_{water} (∼1) and TS_{water} (∼8 kcal/mol). The relatively large statistical errors in both the energy and entropy cancel each other to a large extent in F^{TI}_{water}.

While we argue that ΔF_{loop} should be taken into account (1 and 1′ in Figure 1) the present result, 25 ± 7 kcal/mol was obtained with a large error due to the uncertainty in the experimental conformation of the open loop in apo avidin. However, this is not an inherent limitation, as we have calculated ΔF_{loop} for L3,4 of streptavidin (where both conformations are known) obtaining a comparable result, 27.2 ± 2 kcal/mol with a much smaller error.[45] It should also be pointed out that Lazaridis et al. (using implicit solvation) have estimated the effect of the structural changes in avidin upon its binding of biotin to find an energy of ∼21 kcal/mol, where most of the conformational changes are due to the loop.[46] $V_a\Omega_a$ was estimated from the 2 ns MD trajectory of biotin in the protein environment; we find, $V_a = \sim 244$ Å^3 and $\Omega_a = \sim 8\pi^2/32$, which leads to $k_BT \ln (8\pi^2V^0/\Omega_aV_a) = 3.2 \pm 0.5$ kcal/mol. As pointed out earlier, this calculation is approximate and one can argue that the result, $k_BT \ln (\Omega_aV_a) = \sim 3.8$ kcal/mol, is still an overestimation because Ω_a and V_a were evaluated separately, that is the correlation between them has been ignored. In any case, even 3.8 is much smaller than the error of ∼7 kcal/mol.

The correct evaluation of the entropy due to the change of the external coordinates of biotin in the active site can be obtained, for example, by adding two initial reconstruction steps to those based on the internal coordinates. First, a small cube with volume V is defined around the position $\mathbf{x}$ (relative to the protein) of the external atom of biotin conformation i; the system is then simulated, and the probability density, $p_{\mathbf{x}}(i) = m/(n_fV)$ is calculated where m is the number of snapshots for which the cube was visited by the external atom out of the total n_f snapshots. Then, the external atom is fixed at $\mathbf{x}$ and a similar procedure is carried out for calculating the probability density, $p_{rotation}(i)$ of the external rotation defined by the three Euler angles; thus, $S_{external} = -k_BT \ln [p_{\mathbf{x}}(i)p_{rotation}(i)]$ defines the contribution of the external coordinates to the entropy of the ligand in the active site; this is carried out for each of the n_s members of the sample.

The averaged results (over $n_s = 88$) for $E_{btn-btn}$, $E_{btn-tmpl}$, TS_{btn}, F^{TI}_{water}, F^{TI}_{loop}, and their sum, F_{sum} [and $k_BT \ln (8\pi^2V^0/\ln \Omega_aV_a)$] are presented in Table 3 for the protein and solvent environments; the table also shows the differences Δ between the two sets. The most striking observation is that all the components have small statistical errors. The only exception is the results for F^{TI}_{loop}, which in this study have been obtained with high uncertainty as discussed in the previous paragraph; therefore we present the error (±7 kcal/mol) only for the difference, ΔF_{loop}. This relatively large error affects also the result of the total free energy difference, $\Delta F_{sum} = \Delta A^0 = -24.4 \pm 7$ kcal/mol, which covers the experimental value, −20.7 kcal/mol. Notice, however, that the error in $F_{sum} - F^{TI}_{loop}$, that is, the error in the free energy without the error of F^{TI}_{loop}, is small, 0.5, 0.6, and 0.8 kcal/mol for the protein, solvent, and their difference, respectively; in this calculation the error of $k_BT \ln (8\pi^2V^0/\Omega_aV_a)$ was not considered as well.

To check further the reliability of our results we have calculated ΔA^0 using DDM, by imposing a translational harmonic potential with $K = 3$ kcal/(Å^2) on biotin in the active site;[47] in this case, $\Delta A^0 = \Delta A^0_I - \Delta A^{*,0}_{II} - \Delta A^0_r$ (eqs 29−31 of ref 5). By decoupling biotin from solvent (water) we obtained $\Delta A^0_I = 153.4$ kcal/mol (eq 29, ref 5). For biotin in the active site, the harmonic restraints were imposed first followed by the annihilation of the biotin−environment interactions (water and avidin), which has led to $\Delta A^{*,0}_{II} = 204.2$ kcal/mol (eq 30, ref 5). The final (analytical) release of the restraints leads to $\Delta A^0_r = -k_BT\{\ln [V^0K/(2\pi k_BT)]^{3/2}\} = -4.24$ kcal/mol (eq 31, ref 5). Thus, $\Delta A^0 = 153.4 - 204.2 + 4.2 = -46.6 \pm 3$ kcal/mol. This value is equal within the error bars to the HSMD-TI value, $\Delta A^0 = -50.1 \pm 1.5$ kcal/mol (Table 3 and the following discussion); however, because we have argued that $k_BT \ln (\Omega_aV_a) = \sim 3.8$ kcal/mol is an overestimation, the agreement between the two methods is even better because the central value of the HSMD-TI, $\Delta A^0 = -50.1$ is actually larger becoming closer to the DDM result. Notice that in this comparison the effect of the loop has been ignored, as this effect is not calculated by HSMD-TI but by a standard TI procedure.

IV. Summary and Conclusions

This work is the first application of HSMD-TI for calculating the absolute binding free energy of a ligand to a protein. The avidin−biotin complex is the largest system (2000 template atoms and 350 waters) studied thus far by the method, and the high accuracy achieved is very encouraging. The errors in F^{TI}_{water} are small (partially) because the biotin structures are held fixed

TABLE 3: Energetic, Entropic, and Free Energy Components (in kcal/mol) that Contribute to the Absolute Free Energy of Binding, ΔA^0 (eq 3) Obtained for the Protein and the Solvent[a]

	$k_BT \ln[(8\pi^2/\Omega_a)(V^0/V_a)]$	$E_{btn-btn}$	$E_{btn-tmpl}$	$-TS_{btn}$	F^{TI}_{water}	F^{TI}_{loop}	F_{sum}
protein	3.2 ± 0.5	−85.0 ± 0.5	−72.9 ± 0.5	−41.4 ± 0.2	−6.7 ± 0.2	−130.0	−332.8 ± 0.5
solvent		−85.1 ± 0.4		−48.5 ± 0.3	−19.1 ± 0.4	−155.2	−307.9 ± 0.6
Δ, prot − solv	3.2 ± 0.5	0.1 ± 0.4	−72.9 ± 0.5	7.1 ± 0.6	12.4 ± 0.4	25.2 ± 7	−24.9 ± 7

[a] Results for TS_{btn} and F^{TI}_{water} are taken from Tables 1 and 2, respectively. F_{sum}, the total free energy, is defined up to an additive constant which is expected to be the same for the two environments. The errors of F^{TI}_{loop} are not given and are not included in those for F_{sum}; the large error of ΔF_{loop} is included in $\Delta A^0 = \Delta F_{sum}$.

TABLE 4: Comparison of Energy and Free Energy Results (in kcal/mol) Obtained in the Protein Environment for a Fixed and a Moving Template[a]

template	$E_{btn-btn}$	$E_{btn-bmpl}$	F^{TI}_{water}(ch)	F^{TI}_{water}(LJ)	F_{total}
fixed	−85.0 ± 0.5	−72.9 ± 0.5	−10.3 ± 0.2	3.6 ± 0.1	−164.6
moving	−86.9 ± 0.5	+0.8 ± 0.5	−39.7 ± 0.4	−31.7 ± 0.4	−157.5
difference (Δ)	1.9 ± 1.0	−73.7 ± 1.0	29.4 ± 0.6	35.3 ± 0.5	−7.1 ± 3.0

[a] Both sets of results are based on samples of 88 avidin−biotin configurations; the fixed results are taken from Table 3. The F^{TI}_{water} results for the moving template include integration over both the moving template and water. F_{total} is the sum of the energies and free energies in each row. The errors in the differences (bottom row) are the sum of the errors of the two components.

during TI. This should be distinguished from the commonly used methods where the ligand moves during TI and can leave the active site or alternatively should be kept there by restraints.

An important advantage of the reconstruction process of HSMD is that it *directly* provides the entropy of biotin in the protein and solvent environments hence their difference ΔS_{btn} is known, which is a useful property in rational drug design. In previous studies ΔS was calculated for a loop (or a peptide) populating two different *restricted* regions in conformational space (microstates) where the resulting ΔS is thus small ($\leq$2 kcal/mol) and the required reconstruction runs are relatively short, of 5−10 ps. Here, for the first time the entropies are calculated for a flexible and a highly restricted microstate, where $T\Delta S_{btn} = 7.1 \pm 0.6$ kcal/mol is relatively large for such a small molecule; however, this high accuracy requires longer reconstruction runs of 48−60 ps. (Still, almost the same accuracy has been obtained for $n_s = 40$.) Also, HSMD-TI provides *directly* the change in the entropy of water, TS_{water}, which is again due to the fixed ligand (biotin) during TI.

To the best of our knowledge, this is the first study where ΔF_{loop}, is taken into account in the calculation of the free energy of binding. However, while the accuracy of each (loop) integration is high the uncertainty in the experimental conformation of the open loop in apo avidin leads to a significant error. As pointed out earlier, this is not an inherent limitation as accurate results for ΔF_{loop} were already obtained for streptavidin[45] and other loops,[19−21] where the open conformation is known. The relatively large contribution, $\Delta F_{loop} = \sim$25 kcal/mol suggests that if structural changes in a mobile loop (or other parts of the protein) occur due to ligand binding this effect should be considered.[46,15]

The fact that our result for the absolute free energy of binding, −24.4 ± 7 covers the experimental value, −20.7 kcal/mol is very encouraging. However, our model is based on a fixed template and biotin is treated here uncharged in contrast to the experimental conditions. To check the effect of the fixed template, we allowed a relatively small part of the template to move, carried out two 2 ns MD runs of the new avidin−biotin system from which a sample of 88 configurations has been extracted as before, for TI. The moving part of the template was determined by calculating the distance of each protein atom from the center of biotin; if the distance is smaller than 10 Å the entire respective residue is included in the moving template. In this way, we defined a partial template of 323 atoms (out of 2028), which were moved in the MD simulations while the positions of the rest of the template's atoms were held fixed; the water sphere was defined as before. Notice however that in the TI process the interactions of biotin with water and with the moving template are decreased simultaneously to zero (rather than only the biotin−water interactions as in the case of the entirely fixed template).

The results compared with those of a fixed template appear in Table 4. The table reveals that $E_{btn-btn}$ is almost the same for the two models, whereas $E_{btn-tmpl}$ (for the fixed part of the template) has been changed dramatically from −72 (entire template is fixed) to 0.8 kcal/mol (the outer part of the template is fixed), meaning that beyond a radius of ∼10 Å the *direct* energetic effect of the template is small. Altogether, the moving template leads to an increase of $\Delta F_{total} = 7.1$ kcal/mol in the total free energy. However, one would expect the entropy of biotin S_{btn}(protein) (which was not calculated) to increase as well perhaps by 1−2 kcal/mol thereby decreasing the 7.1 value to 6 or 5 kcal/mol. This suggests that increasing the moving template further will not decrease ΔF_{total} significantly and the effect of the loop is thus imperative for obtaining the experimental result. This calculation also shows that HSMD-TI can easily handle a moving template where the statistical errors are relatively small (Table 4). The effect of charged biotin is unclear as Wang et al. (using LIE)[48] obtained similar results for charged and neutral biotin; we shall study the charge effect by incorporating HSMD-TI within models that take into account long-range electrostatic effects.[49,50]

In summary. We have demonstrated the unique features of HSMD-TI (e.g., the direct calculation of S_{btn} and S_{water}) and have verified that for a simplified model of the biotin−avidin complex the various thermodynamic components are computationally sound. This encouraging study constitutes the basis for the application of HSMD-TI to more realistic models of complexes of biological and therapeutic interest.

Acknowledgment. This work was supported by NIH grant 2-R01 GM066090.

References and Notes

(1) Hermans, J.; Shankar, S. *Isr. J. Chem.* **1986**, *27*, 225–227.
(2) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742–3746.
(3) Miyamoto, S.; Kollman, P. A. *Proteins* **1993**, *16*, 226–245.

(4) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
(5) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.
(6) Zhou, H.-X.; Gilson, M. K. *Chem. Rev.* **2009**, *109*, 4092–4107.
(7) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118–1134.
(8) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. *J. Chem. Phys.* **2005**, *123*, 084108.
(9) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.
(10) Singh, N.; Warshel, A. *Proteins* **2010**, *78*, 1724–1735.
(11) Singh, N.; Warshel, A. *Proteins* **2010**, *78*, 1705–1723.
(12) Fujitani, H.; Tanida, Y.; Matsuura, A. *Phys. Rev. E* **2009**, *79*, 021914.
(13) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. *J. Chem. Phys.* **2006**, *125*, 084901.
(14) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.
(15) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 1231–1235.
(16) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.
(17) Chen, P.; Kuyucak, S. *Biophys. J.* **2009**, *96*, 2577–2588.
(18) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2004**, *121*, 10889–10904.
(19) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Theory Comp.* **2008**, *4*, 192–208.
(20) Cheluvaraja, S.; Mihailescu, M.; Meirovitch, H. *J. Phys. Chem. B* **2008**, *112*, 9512–9522.
(21) Mihailescu, M.; Meirovitch, H. *J. Phys. Chem. B* **2009**, *113*, 7950–7964.
(22) Meirovitch, H. *J. Mol. Recognit.* **2010**, *23*, 153–172.
(23) Green, N. M. *Adv. Protein Chem.* **1975**, *29*, 85–133.
(24) Wilchek, M.; Bayer, E. A. *Trends Biochem. Sci.* **1989**, *14*, 408–412.
(25) Green, N. M. *Avidin and Streptavidin, in Methods in Enzymology*; Wilchek, M., Bayer, E. A., Eds.; Academic Press: London, 1990; *184*, 51−67.
(26) Bayer, E. A.; Wilcheck, M. *Methods Biochem. Anal.* **1980**, *26*, 1–45.
(27) Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M. *J. Mol. Biol.* **1993**, *231*, 698–710.
(28) Livnah, O.; Bayer, E. A.; Wilchek, M.; Sussman, J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5076–5080.
(29) Chu, V.; Freitag, S.; Trong, I. L.; Stenkamp, R. E.; Stayton, P. S. *Protein Sci.* **1998**, *7*, 848–859.
(30) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568–1581.
(31) Pugliese, L.; Malcovati, M.; Coda, A.; Bolognesi, M. *J. Mol. Biol.* **1994**, *235*, 42–46.
(32) Nardone, E.; Rosano, C.; Santambrogio, P.; Curnis, F.; Corti, A.; Magni, F.; Siccardi, A. G.; Paganelli, G.; Losso, R.; Apreda, B.; Bolognesi, M.; Sidoli, A.; Arosio, P. *Eur. J. Biochem.* **1998**, *256*, 453–460.
(33) Repo, S.; Paldanius, T. A.; Hytönen, V. P.; Nyholm, T. K. M.; Halling, K. K.; Huuskonen, J.; Pentikäinen, O. T.; Rissanen, K.; Peter Slotte, J.; Airenne, T. T.; Salminen, T. A.; Kulomaa, M. S.; Johnson, M. S. *Chem. Biol.* **2006**, *13*, 1029–1039.
(34) Freitag, S.; Le Trong, I.; Klumb, L.; Stayton, P. S.; Stenkamp, R. E. *Protein Sci.* **1997**, *6*, 1157–1166.
(35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(36) Ponder, J. W. *TINKER - Software Tools for Molecular Design, Version 5.0*. Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine: St. Louis, Mo, 2009.
(37) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(38) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*: Clarenden Press: Oxford, 1987.
(39) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459–466.
(40) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.
(41) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
(42) Li, Z.; Scheraga, H. A. *Proc Nat. Acad. Sci. USA* **1987**, *84*, 6611–6615.
(43) Nocedal, J. *Math Comp.* **1980**, *35*, 773–782.
(44) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025–9031.
(45) General, I. J.; Meirovitch, H. *J. Chem. Phys.* In press.
(46) Lazaridis, T.; Masunov, A.; Gandolfo, F. *Proteins* **2002**, *47*, 194–208.
(47) Roux, B.; Nina, M.; Pomés, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670–681.
(48) Wang, J.; Dixon, R.; Kollman, P. A. *Proteins* **1999**, *34*, 69–81.
(49) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161–185.
(50) Im, W.; Bernèche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924–2937.

JP1076752