THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

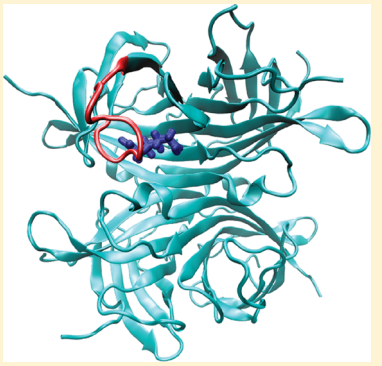
pubs.acs.org/JPCB

Absolute Free Energy of Binding of Avidin/Biotin, Revisited

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

ABSTRACT: The binding of biotin to avidin is one of the strongest in nature with absolute free energy of binding, $\Delta A^0 = -20.4$ kcal/mol. Therefore, this complex became a target for a large number of computational studies, which all, however, are based on approximate techniques or simplified models and have led to a wide range of results Therefore, ΔA^0 is calculated here by rigorous statistical mechanical methods and models that consider long-range electrostatics. (1) We apply our method, "hypothetical scanning molecular dynamics with thermodynamic integration" (HSMD-TI) to avidin−biotin modeled by periodic boundary conditions with particle mesh ewald (PME). (2) We apply the double decoupling method (DDM) to this system modeled by the spherical solvent boundary potential (SSBP) and the generalized solvent boundary potential (GSBP). The corresponding results for neutral biotin, $\Delta A^0 = -29.1 \pm 0.8$ and -25.2 ± 0.5 kcal/mol are significantly lower than the experimental value; we also provide the result for a charged biotin, $\Delta A^0 = -33.3 \pm 0.8$ kcal/mol. It is plausible to suggest that this disagreement with the experiment may stem from ignoring the (positive) contribution of a mobile loop that changes its structure upon ligand binding.



I. INTRODUCTION

Hen egg-white avidin is a tetrameric glycoprotein composed of 4 × 128 residues, which can bind up to four biotin molecules. This binding is one of the strongest in nature, with an absolute (standard) free energy of binding, $\Delta A^0 = -20.4$ kcal/mol ($K_d = 10^{-15}$ M),[1] a property that has been exploited to devise powerful tools for affinity chromatography, biochemical assays, and many other applications.[2−4] Biotin binds to the end of the β-barrel formed by each subunit of avidin as is shown by the crystal structures of the complex. Also the binding of biotin involves the 12-residue mobile loop, 35−46, denoted L3,4 (connecting strands 3 and 4; see Figure 1)[5,6] that changes its conformation from "free" in unbound (apo) avidin to "bound" in the complex.[6−9] The bound conformation which closes on biotin 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[10] (for a detailed discussion see ref 11).

To elucidate the origin of this strong binding, the avidin−biotin complex has become a target for a large number of theoretical−computational studies. However, while reliable models and rigorous techniques for calculating ΔA^0 are available,[12−34] these studies are all based on simplified models or approximate techniques, which have led to a wide range of results for ΔA^0. For example, using LIE, Wang et al.[35] obtained $\Delta A^0 = -21.82$ and -19.65 while Kuhn and Kollman[36] using MM/PBSA obtained −17.7 kcal/mol. $\Delta A^0 = -27.3$ and -14.4 kcal/mol were obtained by Ryde's group from MM/GBSA[37] and MM/3D-RISM theory,[38] respectively, while Tong et al. using PBSA obtained, −23.4 kcal/mol.[39] Therefore, as a first step in elucidating the factors that play a role in the binding of avidin−biotin, it is necessary to calculate ΔA^0 of this important complex with techniques that consist of rigorous statistical mechanics applied to reliable models that take into account

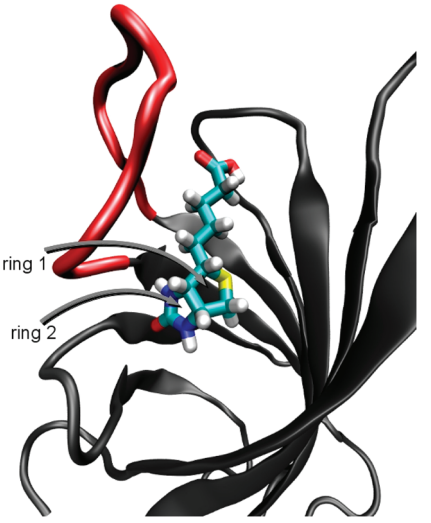


Figure 1. Illustration of one chain of avidin with biotin bound at the active site, which is closed by the mobile loop L3,4 (colored red). The two rings of biotin are numbered; ring 2 is positioned at the bottom of the active site.

Special Issue: Harold A. Scheraga Festschrift

Received: October 13, 2011
Revised: January 25, 2012
Published: February 2, 2012

dx.doi.org/10.1021/jp212276m | J. Phys. Chem. B 2012, 116, 6628−6636

long-range electrostatics. Note also that none of the above studies addressed the loop issue, and no attempts were made to calculate the free energy due to the transition of L3,4 from the free to the bound conformation, $\Delta F_{loop} = F_{bound} - F_{free}$, which in principle should be included in the calculation of ΔA^0. The effect of mobile loops was ignored also in studies of other complexes where such loops are operative. The best way to address this point would be to perform an explicit evaluation of the loop's free energy contribution, but this is not straightforward since the full experimental conformation of the open loop is not known. Therefore, one hopes that a rigorous calculation of ΔA^0 as stated above will shed light on the effect of L3,4 on the binding of avidin−biotin.

Before describing the specific methods used, it should be pointed out that in a standard *rigorous* approach for calculating ΔA^0 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. Two main methods in this category are "the double annihilation method" (DAM)[13,15,16,19,23,24] and the double decoupling method (DDM).[12,14−16,18,20,25,26,30,31] To overcome the end point problem encountered with DAM (i.e., the possible escape of the ligand from the active site at the end of the simulation), with DDM restraints are rigorously added to hold the ligand there. Both methods have been developed systematically in the last 15 years, where various implementation issues have been improved and a large number of complexes have been successfully studied (see the review by Deng and Roux[20]).

In the last several years we have developed a new statistical mechanics method for calculating the absolute free energy and entropy, called hypothetical scanning molecular dynamics [HSMD or HSMC where Monte Carlo replaces molecular dynamics (MD)]; HSMC(D) was tested successfully on systems of increasing complexity.[40−47] While this is a general technique which can be applied to any system (including liquids) we have found that the contribution of solvent to the free energy is calculated more efficiently by a TI procedure that is thus incorporated within the framework of HSMD. The combined method, HSMD-TI was applied to mobile loops in proteins of increasing size and very recently has been extended for calculating protein−ligand affinities by treating first the avidin−biotin complex[11] and later the binding of the ligand L8 to the protein FKBP12.[48]

The main objective of our initial biotin−avidin study[11] was to examine various computational aspects of HSMD-TI, which therefore was applied to a simplified model of the complex. Thus, only part of the protein (the template) around the active site was considered, and this part was covered by a sphere of TIP3P water.[49] Furthermore, only the ligand and water were allowed to move in the MD simulations while the atoms of the template were kept fixed at their crystal structure positions. By not applying cutoffs on nonbonded interactions, we assumed that long-range electrostatic effects were satisfied to a large extent; also we attempted for the first time to calculate ΔF_{loop}. However, the calculation was very approximate due to the simplified model used and the fact that the positions of most of the L3,4 atoms in the crystal structure of apo avidin are not resolved in the crystal structure.[8,9] We obtained $\Delta A^0 = -50.1$ and only after adding $\Delta F_{loop} = 25.2 \pm 7$ the final result, -24.9 ± 7 "covered" the experimental value −20.4 kcal/mol; clearly an uncertainty of 7 kcal/mol although a consequence of the unavailable apo-avidin experimental structure, is unacceptable for calculating ΔA^0. Finally allowing the template to move did not change this picture significantly.

In a following project, the binding free energy of FKBP12-L8 was studied, and since FKBP12 is a small protein (107 residues) and L8 is hydrophobic, we had reasons to believe that applying the modeling used for avidin−biotin would prove effective. However, even though the whole protein was covered with water and all its atoms were allowed to move in the simulations, the results for ΔA^0 were found to be at least 5 kcal/mol above the experimental value, and since this unsatisfactory result could not be "blamed" on a mobile loop, we concluded that one should consider long-range electrostatics more rigorously.[48] Thus, using the AMBER99 force field[50,51] and periodic boundary conditions implemented via particle mesh ewald (PME)[52] has led to $\Delta A^0 = -10.7 \pm 1.0$, in excellent agreement with the experimental values $\Delta A^0 = -10.9$ and −10.6 kcal/mol. We have also applied HSMD-TI to FKBP12-L8 described by the spherical solvent boundary potential (SSBP[53]) and the generalized solvent boundary potential (GSBP[54]) methods, which are implemented in the software CHARMM.[55] With SSBP/GSBP only part of the protein and water around the active site are considered, but long-range electrostatic effects are taken into account. We also treated this model by DDM using restraints and the two methods have led to comparable results for ΔA^0.

These FKBP12-L8 results demonstrate (1) that HSMD-TI constitutes a reliable practical tool for calculating ΔA^0 and related entropies and (2) that taking into account long-range electrostatics is mandatory.

Therefore, in this work we have decided to apply HSMD-TI again to the biotin−avidin complex which is modeled as before by the AMBER99 force field, while periodic boundary conditions with PME are now used. For comparison, ΔA^0 is also calculated by DDM using the GSBP/SSBP model. It should be pointed out that in a previous study of a mobile loop of choline esterase,[45] HSMD-TI results for ΔF_{loop} were found to be in a very good agreement with the experiment; therefore, an issue is whether to attempt calculating ΔF_{loop}(L3,4) again. However, in contrast to the avidin case, in ref 45 (and in our other loop studies[43,44,46]), the X-ray structures of the bound and free loop were well-defined. From our experience, until a better crystal structure for apo avidin becomes available, an improved estimation of ΔF_{loop} will be unfeasible. Therefore, ΔF_{loop} will not be recalculated, but its effect may be deduced indirectly, from the results for ΔA^0 of the present study.

In this context it should also be pointed out that in neutral pH, biotin is expected to be in the ionic state. However, in previous studies biotin was treated both with and without charge. For example, Wang et al.[35] using LIE obtained comparable results for ΔA^0 for a neutral and charged biotin. A neutral biotin was also used by Izrailev et al.[56] and perhaps by Lazaridis et al.[57] (it is not specified). On the other hand, charged biotin was used by Kuhn and Kollman[36] and in several studies of the Ryde's group.[37,38] Therefore, we believe it is of interest to study the avidin−biotin complex with both neutral and charged biotin. Since in ref 11 biotin was treated as neutral, we study neutral biotin first, as comparing the old and new results for ΔA^0 would demonstrate the effect of long-range electrostatics; therefore, the main calculations of this paper are carried with neutral biotin.

It should be pointed out that treating a charged biotin with rigorous statistical mechanics methods might not be computationally straightforward. In particular, our attempts to simulate

a charged biotin using GSBP/SSBP have failed. The entrance of a ligand to the active site of avidin is blocked by the L3,4 loop, therefore application of methods which are based on potential of mean force[34] would not be easy (in fact, in the computer experiments of Izrailev et al.[56] the loop disturbed the pulling of biotin from the active site and had to be removed). While a charged ligand can be treated by models that consider the reaction field (e.g., see ref 58 and references cited therein), we are not aware of their use for calculating ΔA^0. Moreover, for a comparison with the results for neutral biotin, one would seek to investigate a charged biotin within the framework of PME. However, the only DAM/DDM/PME study involving a charged ligand known to us is that of Dixit and Chipot[59] applied to the streptavidin–biotin complex; in these calculations an Na ion was added to both the solvent and the protein systems and its charge was eliminated together with that of biotin. This method is being tested by us in another project where the effect of salt on ΔA^0 is carefully studied. Therefore, to widen the picture, we also provide a HSMD-TI result for ΔA^0 obtained for a charged biotin with the procedure used by Dixit–Chipot.

II. THEORY AND METHODOLOGY

II.1. Theory of Binding. Imagine a *dilute* solution of a protein (P) and a ligand (L) in a volume V in equilibrium with their complex (PL), P + L ↔ PL. The absolute (standard) free energy of binding, ΔA^0 (in the NVT ensemble) is[15,16]

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

where T is the absolute temperature, k_B the Boltzmann constant, and N stands for the number of solvent molecules. $Z = V8\pi^2\bar{Z}$, where $\bar{Z}_{PL,N}$, $\bar{Z}_{P,N}$, 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*, where the integration ($V8\pi^2$) over the external coordinates (e.g., a reference atom and three Euler angles defined by two more atoms) has already been carried out, and V^0 = 1660 Å^3 is the standard volume. Notice that the ligand moves in the active site, i.e., $\bar{Z}_{PL,N}$ includes a localized ligand partition function whose coordinates can also be divided into internal and external, and the contribution of the latter will be calculated by HSMD (rather than analytically as in solvent). ΔA^0 is expressed in terms of configurational (Helmholtz) free energies, $F = -k_B T \ln \bar{Z}$($F_{0,N} = k_B T \ln Z_{0,N}$) and an additional term

$$\Delta A^0 = (F_{PL,N} - F_{P,N}) - (F_{L,N} - F_{0,N}) + k_B T \ln(8\pi^2 V^0) = \\ = \Delta F_p - \Delta F_{sol} + k_B T \ln(8\pi^2 V^0) \quad (2)$$

ΔF_p and ΔF_{sol} are free energy differences defined for the protein and solvent environments, respectively, which are calculated by HSMD-TI. Also, the absolute Gibbs free energy $\Delta G^0 \sim \Delta A^0$ since $\Delta G^0 = \Delta A^0 + P^0 \Delta \bar{V}_{PL}$ where $P^0 \Delta \bar{V}_{PL}$ is small and can be neglected.[15,16]

The complete theory of HSMD-TI applied to binding is presented in a recent paper[48] (see also refs 40–47); therefore, we present here mostly the practical implementation of the method. The process starts by carrying out two production MD runs: one of the ligand in water and another of the protein–ligand complex in water. From these trajectories we determine two sets of n_s equally spaced frames (snapshots), for a later analysis by a reconstruction procedure and TI calculations. We describe first the TI procedure.

II.2. HSMD-TI. For simplicity we start with the system of the ligand in solvent. For each of the related n_s frames (mentioned above), the ligand-solvent interactions are eliminated gradually by a TI procedure, where the electrostatic interactions are eliminated first (using a parameter λ: 1→0) followed by the elimination of the Lennard-Jones interactions using a soft-core potential. These TI results for the n_s frames are averaged and the value of the average constitutes the water contribution to ΔF_{sol} (eq 2). Exactly the same procedure is applied to each of the n_s frames of the complex which leads to the TI contribution to ΔF_p (eq 2). Notice that with HSMD-TI the elimination of the ligand-environment interactions is performed for the *fixed* ligand structure defined by each of the frames, while the other atoms of the system (water and protein) move in the MD simulations; this differs from the DAM and DDM methods where the ligand moves as well.[48] Furthermore, the end-point problem encountered with DAM does not exist and the need for applying restraints is avoided.

II.3. Reconstruction of the Ligand's Internal Entropy. ΔF_p and ΔF_{sol} also have a contribution from the ligand entropy which, with HSMD, is explicitly calculated. This calculation is carried out by a reconstruction procedure whose main features are described below (see also refs 46 and 48) for a ligand in solvent. Thus, for frame i, $1 \le i \le n_s$ one seeks to calculate the Boltzmann probability density P_i^B related to the ligand, as a product of transition probabilities (TPs) (also called conditional probabilities) defined in terms of internal coordinates. First, one determines three reference atoms of the ligand which are kept fixed in their original positions in frame i. The remaining $K/2$ heavy atoms are ordered along the chain and are denoted by k', k' = 1 ... $K/2$. The related K dihedral and bond angles are denoted [α_k], k = 1 ... K. Since bond stretching is ignored,[60–62,48] the position of atom k' is determined solely by the preceding dihedral and bond angles, α_{k-1} and α_k ($k = 2k'$), respectively. We then calculate the variability range,

$$\Delta\alpha_k = \alpha_k(\max) - \alpha_k(\min) \quad (3)$$

where α_k(max) and α_k(min) are the maximum and minimum values of α_k found in each sample.

The TPs are calculated atom-by-atom in $K/2$ steps, where during the process we distinguish between "past" and "future" atoms. Past atoms at step k' are atoms 1 ... $k' - 1$ whose TPs have already been determined; the TPs of the future atoms, k', $k' + 1$, ..., $K/2$ should still be determined. To calculate the TP density $\rho(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1)$ of atom k' ($k = 2k'$), we carry out an MD run, where the future atoms are allowed to move together with all the water molecules, while the past atoms are kept fixed at their positions at frame i. By considering a future conformation (of atoms k', $k' + 1$, ..., $K/2$) every 20 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,[11,43–48]

$$\rho_{\text{ligand}}(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1) \approx \rho^{\text{HS}}(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1) = n_{\text{visit}}/[n_f \delta\alpha_{k-1}\delta\alpha_k J] \quad (4)$$

where $\rho^{\text{HS}}(\alpha_{k-1}\alpha_k|\alpha_{k-2}, ..., \alpha_1)$ becomes exact for very large n_f ($n_f \to \infty$) and very small bins ($\delta\alpha_{k-1}$, $\delta\alpha_k \to 0$). J is the Jacobian. The corresponding probability density related to the ligand's conformation is

$$\rho^{\text{HS}}(\alpha_K, ..., \alpha_1) = \rho^{\text{HS}}([\alpha_k]) = \prod_{k=1,2}^{K-1} \rho^{\text{HS}}(\alpha_k\alpha_{k+1}|\alpha_{k-1}, ..., \alpha_1) \quad (5)$$

which defines, $S^{\text{A}}_{\text{ligand}} = -k_{\text{B}}\int_m \rho^{\text{B}}_{\text{ligand}}([\alpha_k])\ \ln\ \rho^{\text{HS}}([\alpha_k])\, Jd[\alpha_k]$—an approximate entropy functional, which is a *rigorous* upper bound for the correct entropy[40] and $\rho^{\text{B}}_{\text{ligand}}([\alpha_k])$ is the Boltzmann probability density. $S^{\text{A}}_{\text{ligand}}$ is estimated by $\bar{S}^{\text{A}}_{\text{ligand}}$ from an MD (Boltzmann) sample of size n_s

$$\bar{S}^{\text{A}}_{\text{ligand}} = -(k_{\text{B}}/n_s)\sum_{t=1}^{n_s} \ln \rho^{\text{HS}}(t) \quad (6)$$

The internal entropy of the ligand in the active site is calculated in the same way, where at step k' the "future" includes atoms k', $k'+1$, ..., $K/2$, the waters as well as the protein atoms which are all moved by MD in the reconstruction process. We denote the entropies of the ligand in the protein and the solvent environments by, $S^{\text{A}}_{\text{ligand}}(\text{p})$ and $S^{\text{A}}_{\text{ligand}}(\text{sol})$, respectively, where our main interest is in their *converged* difference ΔS_{ligand} which is expected to be the *exact* difference within the statistical error,

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

Thus, we calculate $S^{\text{A}}_{\text{ligand}}(\text{sol})$ and $S^{\text{A}}_{\text{ligand}}(\text{p})$ for increasing n_f and decreasing bins, verifying that both entropies decrease as the approximation improves, i.e., both approach the correct values from above (notice again that the commanding parameter is n_f where the bin size should correspond to the given statistics (n_f) and it cannot be decreased independently). Typically, the convergence of $\Delta S^{\text{A}}_{\text{ligand}}$ is much faster than that of the individual entropies, due to cancellation of comparable errors in $S^{\text{A}}_{\text{ligand}}(\text{sol})$ and $S^{\text{A}}_{\text{ligand}}(\text{p})$. Thus, one can obtain ΔS_{ligand} in a desired accuracy, when the changes in the improved values of $\Delta S^{\text{A}}_{\text{ligand}}$ are smaller than a given statistical error.

II.4. Entropy of the External Coordinates. Calculation of the external entropy of the ligand in the active site[48] is done by reconstructing the TPs of the three reference atoms with positions $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$. Thus, one determines three fixed coordinates in the laboratory frame, $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ where the relative position of $\mathbf{x}_1$ is defined by a "dihedral angle", α_1 "bond" angles, α_2, and the distance $r = |\mathbf{x}_1 - \mathbf{y}_3|$ (Figure 2). TP(1) is obtained by defining three bins for these variables which replace the two bins in the denominator of eq 4. Notice that n_{visit} is the number of times the three bins are visited *simultaneously*. TP(2) consists of a dihedral and a bond angle (based on $\mathbf{y}_2$, $\mathbf{y}_3$, $\mathbf{x}_1$, and $\mathbf{x}_2$, where $|\mathbf{x}_1 - \mathbf{x}_2|$ is assumed constant), and TP(3) is based only on a dihedral angle; $S_{\text{external}}(i) = -k_{\text{B}}\ \ln\ \text{TP}(i)$, $i = 1, 3$ and one obtains

$$S_{\text{external}} = S_{\text{external}}(1) + S_{\text{external}}(2) + S_{\text{external}}(3) \quad (8)$$

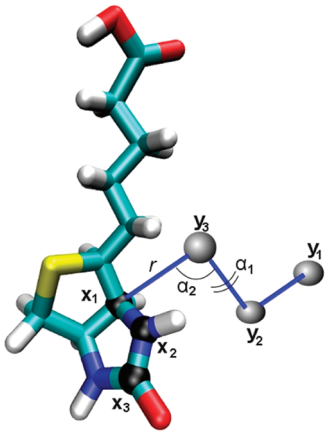


Figure 2. Reconstruction of the external coordinates. $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are the coordinates of three (successive) reference atoms on ring 2 of biotin; $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ are three fixed positions in space. The position $\mathbf{x}_1$ is defined by "dihedral angle" α_1, "bond angle" α_2, and $r = |\mathbf{x}_1 - \mathbf{y}_3|$.

Notice that besides S_{ligand}, one should also include in ΔA^0 (eq 2) the contribution of the intraligand energy, $E_{\text{intraligand}}$, which is averaged over the samples of the two environments (of size n_s) leading to $E_{\text{intraligand}}(\text{sol})$ and $E_{\text{intraligand}}(\text{p})$. With TS_{external}, ΔA^0 (eq 2) becomes

$$\begin{aligned}\Delta A^0 &= \Delta F_{\text{p}} - \Delta F_{\text{sol}} + k_{\text{B}}T\ \ln(8\pi^2 V^0) - TS_{\text{external}} = \\ &= [E_{\text{lintra-ligand}}(\text{p}) - TS_{\text{ligand}}(\text{p}) + F^{\text{TI}}(\text{p})] \\ &\quad - [E_{\text{lintra-ligand}}(\text{sol}) - TS_{\text{ligand}}(\text{sol}) + F^{\text{TI}}(\text{sol})] \\ &\quad + k_{\text{B}}T\ \ln(8\pi^2 V^0) - TS_{\text{external}} = \\ &= \Delta E_{\text{intra-ligand}} - T\Delta S_{\text{ligand}} + \Delta F^{\text{TI}} + k_{\text{B}}T\ \ln(8\pi^2 V^0) \\ &\quad - TS_{\text{external}}\end{aligned} \quad (9)$$

where all the quantities are averages over n_s snapshots and Δ denotes differences in the corresponding variables (protein−solvent); the errors of ΔA^0 and its different components are obtained from the standard deviations (sd) divided by $(n_s)^{1/2}$.

II.5. Calculation of ΔA^0 with GSBP/SSBP. We also calculate ΔA^0 with the GSBP/SSBP[53,54] modeling using DDM. Here the ligand−environment interactions are eliminated by FEP (rather than TI). The equations below which are related to this calculation are based on the notation of ref 16 (eqs 29−31), but to simplify the presentation we define several more notations. Thus, ΔA_{S} (FEP) $\equiv \Delta A^0_{\text{I}}$ (ref 5) and ΔA_{P}(FEP) stands for the free energy change due to the elimination of the ligand−environment interactions in the solvent and the protein, respectively. ΔA_{R} is the free energy due to the application of the restraint, and ΔA^0_r is the free energy due to the release of the restraint plus the effect of the standard state volume. Z is the partition function, $Z = 8\pi^2 V\bar{Z}$ (see eq 1),

dx.doi.org/10.1021/jp212276m | J. Phys. Chem. B 2012, 116, 6628−6636

$$\Delta A_{\mathrm{I}}^{0} = -k_{\mathrm{B}}T \ln \frac{Z_{\mathrm{L},0}Z_{0,N}}{Z_{\mathrm{L},N}} = \Delta A_{\mathrm{S}}(\mathrm{FEP})$$

$$\Delta A_{\mathrm{II}}^{*,0} = -k_{\mathrm{B}}T \ln \frac{Z_{\mathrm{P...L},N}}{Z_{\mathrm{PL},N}} = \Delta A_{\mathrm{R}} + \Delta A_{\mathrm{P}}(\mathrm{FEP})$$

$$\Delta A_{\tau}^{0} = -k_{\mathrm{B}}T \ln \frac{V^{0}Z_{\mathrm{P},N}Z_{\mathrm{L},0}}{VZ_{\mathrm{P...L},N}} = -k_{\mathrm{B}}T \ln \frac{V^{0}K_{r}^{3/2}}{(2\pi k_{\mathrm{B}}T)^{3/2}} \quad (10)$$

where the free energy of binding ΔA^{0} is

$$\Delta A^{0} = \Delta A_{\mathrm{I}}^{0} - \Delta A_{\mathrm{II}}^{*,0} - \Delta A_{\tau}^{0} \quad (11)$$

II.6. Preparation of the Periodic System. We first studied the binding of biotin to avidin using the AMBER10 package[51] with the AMBER99 force field[50] for avidin, the TIP3P model for water,[49] and the GAFF[63] force field (with the AM1-BCC partial charges) for biotin. The unit cell of the periodic system was defined by constructing a truncated octahedron around the tetramer (pdb code: 1avd, ~7800 atoms)[5,7] which contained four biotins; the octahedron was filled with 6920 water molecules. The minimum distance between the protein and the walls of the cell was 8 Å. The system was neutralized with 18 chloride counterions positioned at random locations, while the histidines in the protein were considered neutral. Periodic boundary conditions were implemented via the PME algorithm[52] with a cutoff of 10 Å.

This system was optimized in several stages, starting with a short simulated annealing protocol consisting of energy minimization, where 10^4 steepest descent and 10^4 conjugate gradient steps were performed. Then, using MD, the system was heated to 600 K during 12 ps (using the Berendsen thermostat[64] with a 2 ps time constant), slowly cooled down to 100 K for 28 ps, and cooled down further for 5 ps with a smaller time constant (1.0 ps); finally, the system was driven to 0 K with an even smaller time constant of 0.1 ps for another 5 ps. The time step during this process was 1 fs, and the SHAKE[64] algorithm was applied to all hydrogens.

Next, we performed a temperature and pressure equilibration to 300 K and 1 atm, for 50 ps, using a Berendsen thermostat with a time constant of 1.5 ps and a weak coupling isotropic barostat with a relaxation time of 2 ps; the time step here was 2 fs. Finally, a 2 ns constant volume production run at 300 K (time step of 2 ps) was carried out, where the first 0.4 ns trajectory was used as equilibration. From the latest 1.6 ns, a frame (snapshot) was extracted every 20 ps thus obtaining a sample of 40 frames as potential candidates for future analysis by HSMD-TI.

The solvent system was defined in a similar way where biotin and 482 water molecules were arranged in a truncated octahedron and no counterions were added to this already neutral system. The initial structure of biotin was taken from the crystal structure of the complex, as in ref 11 we have found that the latter structure is similar to that which was obtained from extensive conformational search. Subsequently, an optimization procedure similar to that described above for the complex was applied to the solvent system. 40 frames were determined for this system as well. A check of the movement of biotin in the 2 ns MD trajectories by computer visualization has shown that most of the conformational fluctuations occur in the chain attached to ring 1 (Figure 1); these fluctuations are much more significant in solvent, as seen in Figure 3, where the rmsd

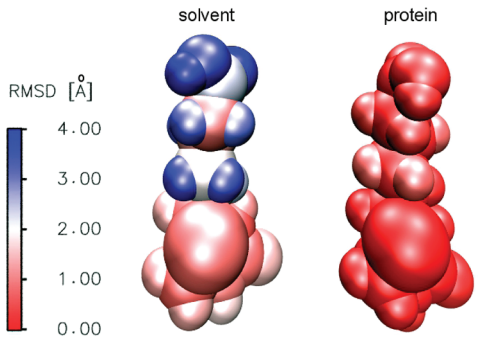


Figure 3. Illustration of the conformational freedom of biotin in the protein and solvent environments obtained from the initial 2 ns trajectories. The rmsd of the atoms in solvent is significantly larger (blue) than in the protein (red).

of the atoms is illustrated by different colors. For biotin in solvent, full rotations are observed at the end of the chain at a relatively large pace, suggesting that the relevant conformational space was sufficiently sampled.

II.7. Calculation of the Entropy. Biotin is a small peptide consisting of two contiguous rings, denoted 1 and 2 (Figure 1); ring 2 is located toward the inner part of the active site, where a linear chain starting from ring 1 points toward the outside of the pocket. Therefore, we defined three atoms in ring 2 as reference atoms (see Figure 2), which allows us to define a physically meaningful external entropy, related to the global movement of biotin in the active site. We first discuss results for the internal entropy. We reconstructed $n_s = 25$ biotin conformations (denoted i) in both, the solvent and the protein, where 17 atoms ($k' = 1, ..., 17$) and 34 angles α_k ($1 \le k \le K = 34$) participate in the reconstruction (see section II.3). Each reconstruction step starts from conformation i with a 120 ps production run and a future biotin conformation is stored every 20 fs for a later analysis; thus, the total sample for each step consists of $n_f = 6000$ future conformations, with the first 200 usually dropped as part of the equilibration. The number of counts, n_{visit} (eq 4), for each pair of bins is calculated leading to $TP_{k'}$, where the product of the 17 TPs is the distribution, ρ^{HS} (eq 5), from which the entropy, S^{A} (eq 6), is calculated.

In practice, the calculation is done in two stages, where in stage 1 we carry out the reconstruction simulations; thus, for the 17 reconstructed atoms, $17n_f$ future chains are generated for snapshot i and their coordinates are stored in a file for a later analysis in stage 2. Stage 1 can be performed in a straightforward way with any of the available programs AMBER, TINKER, CHARMM, etc. In stage 2, the files generated in stage 1 are read by an analysis program, which enables one to calculate the transition probabilities and to study the behavior of S^{A}_{ligand} (eq 6), ΔS^{A}_{ligand} (eq 7), and $S_{external}$ (eq 8) as a function of various parameters (e.g., bin size, n_s, and n_f). This free program with a tutorial and explanations appears at http://www.ccbb.pitt.edu/Faculty/meirovitch/reconstruction-web/reconstruction-web.html

III. RESULTS AND DISCUSSION

III.1. Results for the Internal Entropy. The results for S^{A}_{ligand} (eq 6) are presented in Table 1. As expected by theory, the results for S^{A}_{ligand} decrease systematically as the approximation improves, i.e., with increasing n_p, but they remain almost unchanged as a function of bin size δ; however, these results have not converged even for $n_f = 6000$. On the other

dx.doi.org/10.1021/jp212276m | J. Phys. Chem. B 2012, 116, 6628–6636

Table 1. Results for the Internal Entropy S^{A}_{ligand} of the Ligand (Biotin) in the Solvent and Protein Environment and the Difference $\Delta S^{A}_{ligand} = S^{A}_{ligand}(sol) - S^{A}_{ligand}(p)$ Obtained with AMBER99/PME[a]

bin size, δ	n_f	$TS^{A}_{ligand}(sol)$	$TS^{A}_{ligand}(p)$	$T\Delta S^{A}_{ligand}$
$\Delta\alpha_k/30$	1000	3.2	1.1	2.1
	2000	−3.4	−6.2	2.8
	3000	−7.1	−9.8	2.7
	4000	−9.5	−12.3	2.8
	5000	−11.2	−14.1	2.9
	6000	**-12.7**	**-15.5**	**2.8**
$\Delta\alpha_k/60$	1000	3.0	0.9	2.1
	2000	−3.5	−6.5	3.0
	3000	−7.2	−10.0	2.8
	4000	−9.6	−12.6	3.0
	5000	−11.3	−14.3	3.0
	6000	**-12.8**	**-15.8**	**3.0**
$\Delta\alpha_k/90$	1000	3.1	0.9	2.2
	2000	−3.5	−6.5	3.0
	3000	−7.2	−10.2	3.0
	4000	−9.6	−12.8	3.2
	5000	−11.4	−14.4	3.1
	6000	**-12.9**	**-16.0**	**3.1**
converged				**3.0 ± 0.2**

[a]The results were obtained by reconstructing $n_s = 25$ structures of biotin selected homogeneously from MD samples of 1.6 ns for the solvent and the protein environments. The results are calculated as functions of $\delta = \Delta\alpha_k/l$ and n_f (eqs 3 and 4)—the bin and sample size of the future chains, respectively. S^{A}_{ligand} is defined up to an additive constant (depending on the definition of the angles). The (best) results for $n_f = 6000$ are bold-faced.

hand, the corresponding results for ΔS^{A} show convergence to 3.0 ± 0.2 kcal/mol which is thus considered to be the exact result within the error bars (the same result was also obtained for a smaller sample of $n_s = 20$). The fact that this result for ΔS^{A} is smaller than ΔS^{A} =7 kcal/mol obtained in ref 11 is expected as in ref 11 the protein coordinates were kept fixed during the reconstruction process, leading to a confined biotin in the active site with a relatively low internal entropy.

III.2. Results for the External Entropy. The contributions of the three reference atoms to the external entropy appear in Table 2 as a function of n_f and only a single bin size $\Delta\alpha_k/60$, because the same results (within the statistical errors) were obtained for bin sizes within the range $\Delta\alpha_k/30$−$\Delta\alpha_k/90$. n_f is increased here up to 12000, and as expected, for each atom the entropy decreases as the approximation improves, i.e. as n_f increases but for the larger values of n_f the results show convergence (i.e., they stabilize).

The space covered by these variables can be estimated from eq 4, using the results obtained for the largest bin, $\Delta\alpha_k$ for which $n_{visit}/n_f = 1$, or more specifically by calculating $\exp(TS^{A}/0.6)$, where TS^{A} stands for the results in the table and $k_BT = 0.6$ at 300 K.

III.3. TI Results. To each of the $n_s = 25$ frames of the solvent sample, we applied a TI procedure where the ligand−water interactions were turned off gradually for a *fixed* biotin structure; for the 25 frames of the protein environment, we decoupled both the ligand−water and ligand−protein interactions. Using a parameter λ, $0 \le \lambda \le 1$, the electrostatic interactions were decoupled first followed by decoupling the Lennard-Jones (LJ) potentials (in the presence of zero electrostatic interactions). In all, 30 λ values (windows) were used, 13 for the electrostatic interactions (λ = 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, and 0.99), and 17 for LJ (λ = 0.01, 0.03, 0.07, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.85, 0.90, 0.93. 0.97, and 0.99). In the LJ integration, we used a soft-core potential based on $\delta = 3$ Å^2.[46,65] For a given frame (i), each integration step (window) always started from the initial structure of i according to the corresponding step potential energy $[E(\lambda)]$, followed by a 120 ps production run, where the initial 20 ps are discarded for equilibration.

The results for F^{TI} obtained by eliminating the charge (ch) and the LJ also include intraligand contributions,[66] and because the conformations are fixed, these contributions constitute part of the intraligand interaction energy. We need to remove these contributions from the integration result, since they are already counted in the previous reconstruction step ($\Delta E_{intraligand}$ in eq 9). To calculate these energies, we generated two samples of the ligand in vacuum by removing the protein and water and applied TI to both (vacuum) samples. These vacuum integrations led to −147.9 and −148.0 kcal/mol for the solvent and protein environments, respectively, which were subtracted from the original results for F^{TI}. All the subtracted results appear in Table 3 where $\Delta F^{TI} = F^{TI}(p) - F^{TI}(s) = -40.8$ kcal/mol, which is expected to be a reliable result as in ref 48 we have found that TI of 120 ps leads to convergent results for a much larger ligand than biotin; also, we obtained here $\Delta F^{TI} = -40.9$ kcal/mol based on a smaller sample of $n_s = 10$. The results, $E_{intraligand} = -31.5$ and −29.4 kcal/mol for biotin in

Table 2. External Entropy for the Three Reference Atoms Based on AMBER99/PME[a]

bin size, δ	n_f	$TS^{A}_{ligand}(1)$	$TS^{A}_{ligand}(2)$	$TS^{A}_{ligand}(3)$	$T\Delta S^{A}_{ligand}$(total)
$\Delta\alpha_k/60$	2000	1.8	0.2	0.7	2.7
	4000	1.7	−0.1	0.3	1.9
	6000	1.6	−0.3	0.1	1.4
	8000	1.6	−0.5	−0.1	1.0
	10000	1.5	−0.6	−0.3	0.6
	12000	**1.4**	**-0.6**	**-0.4**	**0.4**
$\Delta\alpha_k$	12000	1.8	−0.25	−0.0	
space covered		20 Å^3	0.65 out of $4\pi = 12.6$ (5.2%)	57°	

[a]The results are based on the 25 structures of biotin used to calculate the internal entropy (Table 1). They are presented for bins sizes $\delta = \Delta\alpha_k/60$ and $\Delta\alpha_k$ (eq 3), and for n_f (eq 4)—the sample size of the future chains with a maximal value of 12000. $S^{A}_{ligand}(1)$ is the entropy related to the volume occupied by the first atom, $S^{A}_{ligand}(2)$ is related to the "bond angle" and "dihedral angle" of atom 2, and $S^{A}_{ligand}(3)$ to the "dihedral angle" of atom 3. In the last row, we provide the space covered by the variables of each atom. The statistical error for $T\Delta S^{A}_{ligand}$(total) (eq 7) is 0.5 kcal/mol. The (best) results for $n_f = 12000$ are bold-faced.

dx.doi.org/10.1021/jp212276m | J. Phys. Chem. B 2012, 116, 6628−6636

Table 3. AMBER99/PME Results for the Free Energy, FTI (kcal/mol) Obtained by TI for the Protein and the Solvent Environments[a]

	F^{TI}(ch)	F^{TI}(LJ)	F^{TI}
protein	−40.9 ± 0.4	−28.7 ± 0.9	−69.6 ± 0.7
solvent	−27.1 ± 0.2	−1.7 ± 0.2	−28.8 ± 0.2
Δ = prot − sol	−13.8 ± 0.4	−27.0 ± 0.5	−40.8 ± 0.5

[a] F^{TI}(ch) and F^{TI}(LJ) are free energies calculated by TI by eliminating the electrostatic and LJ interactions, respectively, based on n_s = 20 structures for each environment. F^{TI} is their sum.

solvent and the protein, respectively, lead to $\Delta E_{intraligand}$ = 2.1 kcal/mol.

Table 4 presents several energetic and entropic components that lead to our estimation of the absolute binding free energy, ΔA^0 = −29.1 ± 0.8 kcal/mol, which is significantly larger (by ∼21 kcal/mol) than ΔA^0 = −50.1 kcal/mol obtained in our ref 11; this change is mainly due to the long-range electrostatic effect provided by PME. Still ΔA^0 = −29.1 is significantly smaller than the experimental value ΔA^0 = −20.4. This gap in the free energy is discussed further in section IV.

The efficiency of HSMD-TI has been discussed in ref 48 and has been estimated to be comparable to that of DDM; more specifically, the efficiency depends on the efficiencies of TI and the reconstruction process where the latter consists purely of MD simulations. Thus, in the protein environment a single TI step (electrostatic or Lennard-Jones) of 120 ps requires ∼15 h CPU on a single processor (Dual Core AMD Opteron, 2.2 GHz); since TI consists of 30 such steps, the total time is 30 × 15 = 450 CPU h. One reconstruction step of 120 ps also requires 15 h CPU which multiplied by the 17 steps (for biotin) leads to ∼255 h, resulting in a total time of ∼705 h. Because the solvent system is smaller, the calculations are ∼4 times shorter, i.e. require ∼176 h. Thus, for n_s = 20, the total time required for calculating ΔA^0 is (176 + 707) × 20 = ∼17600 h CPU.

III.4. SSBP/GSBP Results for Avidin−Biotin Obtained with DDM. Unlike the above model, where the entire complex and surrounding water are considered explicitly, models have been developed where only part of the complex, around the active site is treated explicitly and long-range electrostatics is taken into account implicitly. To this category belong the models of Warshel's group (implemented within the program MOLARIS[67−69]) and SSBP/GSBP mentioned earlier.[20,53,54] We decided to calculate ΔA^0 by applying DDM to the avidin−biotin complex modeled by GSBP and the solvated biotin by SSBP.

The force-field parameters for the ligand were taken from the CHARMM general force field,[70] and the charges of the histidine residues were kept neutral. Most importantly, our DDM procedure is based on a single harmonic distance restraint with a force constant, K_r = 7 kcal/(mol Å^2) (as compared to several different restrains commonly applied within the framework of GSBP/SSBP). However, applying a single restraint is justified because biotin is highly restricted in the active site as discussed earlier.

Using PME and Langevin dynamics at T = 300 K with a friction coefficient of 5 ps^{-1}, we first carried out two 1 ns simulations (1) for the avidin−biotin tetramer (i.e., with four bound biotins) soaked in 8531 TIP3P water molecules and (2) for a single biotin soaked in 3731 water molecules. These initial systems were subsequently reduced by the SSBP/GSBP software as described in the next paragraph. Notice that with this procedure the ligand−environment interactions are eliminated by FEP (rather than TI) in the following order, electrostatic, attractive LJ, and repulsive LJ.

The solvent system (modeled by SSBP), which was "cut" from the larger one described above consists of 762 water molecules within a sphere of ∼18 Å, whereas the protein system (described by GSBP) contains 509 waters within a sphere of ∼20 Å. Each system was simulated by Langevin dynamics at T = 300 K (with a friction coefficient of 5 ps^{-1}) for 4 ns where five initial configurations were selected for FEP calculations from the last 2 ns trajectory (the first 2 ns were used for equilibration). The results for ΔA^0 are calculated with eqs 10 and 11.

The elimination of biotin−water interactions was carried out in 20 windows for each type of interaction (electrostatic, attractive LJ, and repulsive LJ) where for each window the first 30 ps were used for equilibration. The production time for the electrostatic and attractive LJ was 90 ps, which was increased to 120 ps for the repulsive LJ. The different FEP results in Table 5 are averages of the five runs with the errors given by the standard deviation (sd), sd/5$^{1/2}$; we obtain ΔA^0 = −25.2 ± 0.5 kcal/mol. In ref 48, we have found that the GSBP results for the protein system might be sensitive to the density of water, which is difficult to control accurately in this relatively small system. (A grand-canonical procedure could be used to control the water density[71]). Also, perturbation of the repulsive LJ is the most difficult among the three. To check these effects, we applied FEP to a single (protein) frame with additional 20 water molecules and found that the combined FEP results slightly decreased from 45.4 ± 0.4 in Table 5 to 44.9 kcal/mol. For one (protein) frame, we also increased the production run for the repulsive FEP from 120 to 180 ps and found that $\Delta A(FEP)_{LJ\text{-}repulsive}$ = −21.6 ± 0.7 in Table 5 was slightly decreased to −22.5 kcal/mol. These checks suggest that ΔA^0 = −25.2 ± 0.5 kcal/mol in Table 5 might be an upper bound and within the error bars the correct result would not differ much from ΔA^0 = −29.1 ± 0.8 kcal/mol obtained by PME using the AMBER99 force field, i.e. both results are significantly lower than the experimental value −20.6 kcal/mol.

Table 4. Energetic, Entropic, and Free Energy Components (kcal/mol) which Contribute to the Absolute Free Energy of Binding, ΔA^0 (Equation 9) Obtained for the Protein and Solvent with AMBER99/PME[a]

	$k_BT \ln(8\pi^2V^0)$	$-TS_{external}$	$-TS_{ligand}$	$E_{intraligand}$	F^{TI}	total
protein	7.0	−0.4 ± 0.5	15.8. ± 0.5	−29.4 ± 0.3	−69.6 ± 0.7	−76.6 ± 1.6
solvent			12.8 ± 0.6	−31.5 ± 0.6	−28.8 ± 0.2	−47.5 ± 1.3
Δ = prot − solv	7.0	−0.4 ± 0.5	3.0 ± 0.2	2.1 ± 0.7	−40.8 ± 0.5	−29.1 ± 0.8

[a] The table is organized according to eq 9. Results for TS_{ligand}, $TS_{external}$, and F^{TI} are taken from Tables 1, 2, and 3, respectively. Most of the components are defined up to an additive constant and only their difference has a physical meaning. The absolute free energy of binding is ΔA^0 = −29.1 ± 0.8 kcal/mol is defined in the right-hand side of the bottom row.

dx.doi.org/10.1021/jp212276m | J. Phys. Chem. B 2012, 116, 6628−6636

Table 5. Free Energy Results (kcal/mol) Obtained by Applying DDM to the SSBP/GSBP/CHARMM27 Model[54] [a]

	$\Delta A(\mathrm{FEP})_{\mathrm{charge}}$	$\Delta A(\mathrm{FEP})_{\mathrm{LJ\text{-}attractive}}$	$\Delta A(\mathrm{FEP})_{\mathrm{LJ\text{-}repulsive}}$	ΔA_{R}	ΔA^0_τ	total
solvent	21.7. ± 0.1	25.5 ± 0.4	−27.0 ± 0.2			20.2 ± 0.3
protein	27.1. ± 0.4	42.7 ± 0.3	- 21.6 ± 0.7	2.2. ± 0.2	−5.0	45.4 ± 0.4
Δ = solv − prot	−5.4 ± 0.2	−17.2 ± 0.3	−5.4 ± 0.4	−2.2 ± 0.2	5.0	−25.2 ± 0.5

[a]The elimination of the biotin−environment interactions was obtained by free energy perturbation (FEP) in three steps, treating first the electrostatic interactions, then the attractive and repulsive LJ interactions. ΔA_{R} and ΔA^0_τ and $\Delta A^0 = -25.2 \pm 0.5$ kcal/mol are defined in eqs 10 and 11. For each environment, the results are averages of five FEP runs started from different ligand or protein−ligand structures.

IV. CONCLUSIONS

This paper presents the first study of the avidin−biotin complex by rigorous statistical mechanics procedures and models that consider long-range electrostatics. HSMD-TI and DDM are applied to this system modeled by AMBER99/PME and CHARMM27/SSBP/GSBP, respectively. The fact that the corresponding results obtained for ΔA^0 are significantly lower than the experimental value (by ∼9 and 5 kcal/mol) is important as they were obtained by different force fields and methods. These results are expected to be more reliable than those obtained in previous studies using approximate procedures and modeling.[35−39,57] However, as mentioned earlier, under the experimental conditions where the binding affinity was measured, biotin is expected to be negatively charged while in our calculations it has a neutral charge.

Therefore, it is of interest to provide here the HSMD-TI(PME) result, $\Delta A^0 = -33.3 \pm 0.8$ kcal/mol (based on n_s = 20 frames) obtained for a charged biotin in another project, where an Na ion was added to both the solvent and the protein systems and its charge was eliminated together with that of biotin; the components of ΔA^0 are the following: $-TS_{\mathrm{external}}$ = 1.3, $-T\Delta S_{\mathrm{ligand}}$ = 2.9, $\Delta E_{\mathrm{intraligand}}$ = 1.0, and ΔF^{TI} = −45.5 kcal/mol (see eq 9). This result for ΔA^0 is ∼13 kcal/mol below the experimental value, and its closeness to $\Delta A^0 = -29.1 \pm 0.8$ obtained for neutral biotin is theoretically intriguing, as one might wonder whether this proximity demonstrates a general trend. As has already been mentioned, Wang et al.[35] using LIE also obtained comparable results for neutral and charged biotin, ΔA^0 = −21.82 and −19.65 kcal/mol, respectively, which however, are close to the experimental value.

It has long been recognized that conformational changes occurring in the protein upon ligand binding should be taken into account in the calculation of ΔA^0. First attempts to calculate such effects were made by Lazaridis et al.[57] who used an implicit model and found for avidin a reorganization energy of ∼20 kcal/mol which opposes binding; this value is compatible with our approximate estimate in ref 11, $\Delta F_{\mathrm{loop}} = 25 \pm 7$ kcal/mol. They also estimated the entropy change due to the conformational transition of L3,4. (The effect of protein conformational change was also addressed by Mobley et al.[26]). Therefore it is plausible to suggest that the free energy gaps of ∼5, 9, and 13 kcal/mol below the experimental value may stem from ignoring ΔF_{loop} due to L3,4, which exhibits the most obvious conformational change of avidin upon binding.

This picture can be elucidated further by studying complexes with mobile loops where their crystallographic structures are well resolved in both the apo and holo proteins, such as for biotin−streptavidin. However, notice that a direct reliable calculation of ΔF_{loop} without imposing structural restrictions is not trivial since other conformational differences between the apo and holo proteins typically exist as was demonstrated for choline esterase.[72]

AUTHOR INFORMATION

Corresponding Author

*E-mail: hagaim@pitt.edu. Phone: 412-648-3338.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by NIH grant 2-R01 GM066090-4 and by NSF through TeraGrid resources provided by the Lousiana Optical Network Initiative.

REFERENCES

(1) Green, N. M. *Adv. Protein Chem.* **1975**, *29*, 85−133.
(2) Wilchek, M.; Bayer, E. A. *Trends Biochem. Sci.* **1989**, *14*, 408−412.
(3) Green, N. M. *Avidin and streptavidin*. In *Methods in Enzymology*; Wilchek, M., Bayer, E. A., Eds.; Academic Press: London, 1990; Vol. *184*, pp 51−67.
(4) Bayer, E. A.; Wilcheck, M. *Methods Biochem. Anal.* **1980**, *26*, 1−45.
(5) Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M. *J. Mol. Biol.* **1993**, *231*, 698−710.
(6) Livnah, O.; Bayer, E. A.; Wilchek, M.; Sussman, J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5076−5080.
(7) Pugliese, L.; Malcovati, M.; Coda, A.; Bolognesi, M. *J. Mol. Biol.* **1994**, *235*, 42−46.
(8) Nardone, E.; Rosano, C.; Santambrogio, P.; Curnis, F.; Corti, A.; Magni, F.; Siccardi, A. G.; Paganelli, G.; Losso, R.; Apreda, B.; et al. *Eur. J. Biochem.* **1998**, *256*, 453−460.
(9) 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; et al. *Chem. Biol.* **2006**, *13*, 1029−1039.
(10) Chu, V.; Freitag, S.; Trong, I. L.; Stenkamp, R. E.; Stayton, P. S. *Protein Sci.* **1998**, *7*, 848−859.
(11) General, I. J.; Dragomirova, R.; Meirovitch, H. *J. Phys. Chem. B* **2011**, *115*, 168−175.
(12) Hermans, J.; Shankar, S. *Isr. J. Chem.* **1986**, *27*, 225−227.
(13) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742−3746.
(14) Miyamoto, S.; Kollman, P. A. *Proteins* **1993**, *16*, 226−245.
(15) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(16) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.
(17) Zhou, H.-X.; Gilson, M. K. *Chem. Rev.* **2009**, *109*, 4092−4107.
(18) 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.
(19) 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*, 084108−5.
(20) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.
(21) Singh, N.; Warshel, A. *Proteins* **2010**, *78*, 1724−1735.
(22) Singh, N.; Warshel, A. *Proteins* **2010**, *78*, 1705−1723.
(23) Fujitani, H.; Tanida, Y.; Matsuura, A. *Phys. Rev. E* **2009**, *79*, 021914−12.
(24) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. *J. Chem. Phys.* **2006**, *125*, 084901−12.

(25) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683−7689.
(26) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 1231−1235.
(27) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235−10253.
(28) Mobley, D. L.; Dill, K. A. *Structure* **2009**, *17*, 489−498.
(29) Roux, B.; Nina, M.; Pomes, R.; Smith, J.∼C. *Biophys. J.* **1996**, *71*, 670−681.
(30) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707−2714.
(31) Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.
(32) Chen, P.; Kuyucak, S. *Biophys. J.* **2009**, *96*, 2577−2588.
(33) Jiang, W.; Roux, B. *J. Chem. Theory Comput.* **2010**, *6*, 2559−2565.
(34) Woo, H-June; Roux, B. *Proc. Natl. Acad. Sci USA* **2005**, *102*, 6825−6830.
(35) Wang, J.; Dixon, R.; Kollman, P. A. *Proteins* **1999**, *34*, 69−81.
(36) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786−3791.
(37) Genheden, S.; Ryde, U. *J. Comput. Chem.* **2010**, *31*, 837−846.
(38) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. *J. Phys. Chem. B* **2010**, *114*, 8505−8516.
(39) Tong, Y.; Mei, Y.; Li, Y. L.; Ji, C. G.; Zhang, J. Z. H. *J. Am. Chem. Soc.* **2010**, *132*, 5137−5142.
(40) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2004**, *121*, 10889−10904.
(41) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2005**, *123*, 214908−11.
(42) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Phys.* **2005**, *122*, 054903−14.
(43) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Theory Comp.* **2008**, *4*, 192−208.
(44) Cheluvaraja, S.; Mihailescu, M.; Meirovitch, H. *J. Phys. Chem. B* **2008**, *112*, 9512−9522.
(45) Mihailescu, M.; Meirovitch, H. *J. Phys. Chem. B* **2009**, *113*, 7950−7964.
(46) General, I. J.; Meirovitch, H. *J. Chem. Phys.* **2011**, *134*, 025104−17.
(47) Meirovitch, H. *J. Mol. Recognit.* **2010**, *23*, 153−172.
(48) General, I. J.; Dragomirova, R.; Meirovitch, H. *J. Chem. Theory Comp.* **2011**, *7*, 4196−4207.
(49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.. *J. Chem. Phys.* **1983**, *79*, 926−935.
(50) Cornell, W. D.; Cieplak, P.; Bayly, C. L.; 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.
(51) Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R; Walker, R. C.; Zhang, W.; Merz, K. M.;et al. *AMBER 11*, University of California, San Francisco, 2010.
(52) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089−92.
(53) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(54) Im, W.; Bernéche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924−2937.
(55) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dumbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(56) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568−1581.
(57) Lazaridis, T.; Masunov, A.; Gandolfo, F. *Proteins* **2002**, *47*, 194−208.
(58) Donnini, S.; Mark, A. E.; Juffer, A. H.; Villa, A. *J. Comput. Chem.* **2005**, *26*, 115−122.
(59) Dixit, S. B.; Chipot, C. *J. Phys. Chem. A* **2001**, *105*, 9795−9799.
(60) Hnizdo, V.; Darian, E.; Fedorowicz, A.; Demchuk, E.; Li, S.; Singh, H. *J. Comput. Chem.* **2007**, *28*, 655−668.
(61) Killian, B. J.; Kravitz, J. Y.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127*, 024107−16.
(62) Hnizdo, V.; Tan, J.; Killian, B. J.; Gilson, M. K. *J. Comput. Chem.* **2008**, *29*, 1605−1614.
(63) Wang, J.; Wolf, R .M.; Calldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(64) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*: Clarenden Press: Oxford, 1987.
(65) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025−9031.
(66) Steinbrecher, T.; Mobley, D. L.; Case, A. C. *J. Chem. Phys.* **2007**, *127*, 21410813.
(67) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochim. Biophys. Acta* **2006**, *1764*, 1647−1676.
(68) King, E.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647−3661.
(69) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161−185.
(70) Vanommeslaeghe, K; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerel, A. D., Jr. *J. Comput. Chem.* **2010**, *31*, 671−690.
(71) Deng, Y.; Roux, B. *J. Chem. Phys.* **2008**, *128*, 115103−8.
(72) Mihailescu, M.; Meirovitch, H. *Entropy* **2010**, *12*, 1946−1974.