PROTEINS: Structure, Function, and Genetics 47:194–208 (2002)

Contributions to the Binding Free Energy of Ligands to Avidin and Streptavidin

Themis Lazaridis,* Artëm Masunov, and Francois Gandolfo
Department of Chemistry, City College of the City University of New York, New York, New York

ABSTRACT **The free energy of binding of a ligand to a macromolecule is here formally decomposed into the (effective) energy of interaction, reorganization energy of the ligand and the macromolecule, conformational entropy change of the ligand and the macromolecule, and translational and rotational entropy loss of the ligand. Molecular dynamics simulations with implicit solvation are used to evaluate these contributions in the binding of biotin, biotin analogs, and two peptides to avidin and streptavidin. We find that the largest contribution opposing binding is the protein reorganization energy, which is calculated to be from 10 to 30 kcal/mol for the ligands considered here. The ligand reorganization energy is also significant for flexible ligands. The translational/rotational entropy is 4.5–6 kcal/mol at 1 M standard state and room temperature. The calculated binding free energies are in the correct range, but the large statistical uncertainty in the protein reorganization energy precludes precise predictions. For some complexes, the simulations show multiple binding modes, different from the one observed in the crystal structure. This finding is probably due to deficiencies in the force field but may also reflect considerable ligand flexibility. Proteins 2002;47:194–208.** © 2002 Wiley-Liss, Inc.

Key words: association; affinity; reorganization energy; translational/rotational entropy; molecular modeling; docking

INTRODUCTION

Molecular recognition, the specific binding between molecules in solution, is a central element in the workings of a living cell. For example, it is involved in enzyme catalysis, signal transduction, regulation of gene expression, and the mechanism of action of drugs. Molecular recognition is also central in the emerging field of supramolecular chemistry and nanotechnology and in industrial separation processes such as affinity chromatography. Because experimental determination of the binding constants between a large number of pairs of molecules is costly and time-consuming, it would be convenient to obtain information by theoretical and computational means. Given two molecules, it would be useful to be able to predict (a) whether they will bind to each other, (b) how strongly they will bind to each other, and (c) in what configuration they will bind to each other (the docking problem). This work focuses on the second problem: given the structure of a ligand-protein complex, predict the binding free energy from first principles and understand the physical origin of this binding affinity based on theory and computer simulations.

Much of the computational work in the past has been concerned with calculations of relative binding free energies (i.e., prediction of the difference in binding free energy between slightly different inhibitors binding to the same protein or the same inhibitor binding to variants of one protein.) Such calculations are usually performed by using free energy simulations with thermodynamic cycles.[1] Applications include the binding of trimethoprim derivatives to dihydrofolate reductase,[2] binding of substrate to tyrosyl-tRNA synthetase mutants,[3] inhibitors of thermolysin,[4] sulfonamide inhibitors of carbonic anhydrase,[5] benzamidine inhibitors to trypsin,[6] and distamycin analogs to DNA,[7] to mention just a few.

Attempts to predict absolute binding affinities have only recently been reported. Hermans and Wang[8] reported a calculation of the binding free energy of benzene in a cavity of a mutant T4 lysozyme. This was accomplished by combining free energy simulations with restraints and analytical results that incorporate the choice of standard state. A similar methodology was applied to the binding of camphor to cyt P450[9] and of water into the bacteriorhodopsin proton channel.[10] Froloff et al.[11] used continuum electrostatics, together with a surface area model for the nonpolar contribution, to obtain the interaction and desolvation contributions to the binding free energy of peptides to a protein. The calculations were performed on single configurations without any thermal averaging. Because the translational/rotational entropy and the protein reorganization upon binding were neglected, the calculated binding free energies were more negative than the experimental values. The same methodology was applied to trypsin-inhibitor complexes.[12]

Abbreviations: MD, molecular dynamics; LIN, FSHPQNT linear peptide; CYCL, cyclo-CHPQGPPC peptide.

Grant sponsor: Petroleum Research Fund, administered by the American Chemical Society; Grant number: 34680-G4.

Francois Gandolfo's present address is Unilever Research Vlaardingen, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands.

*Correspondence to: Themis Lazaridis, Department of Chemistry, City College of the City University of New York, 138th St. & Convent Ave., New York, NY 10031. E-mail: themis@sci.ccny.cuny.edu

Received 27 August 2001; Accepted 30 November 2001

An alternative method, applicable to flexible ligands, was proposed by Gilson and coworkers.[13] Their approach involves the approximate calculation of configurational integrals for the free and bound states using an implicit solvation model and gives the free energy directly. The entropic contributions are evaluated indirectly by taking the difference between the free energy and the average effective energy (effective energy = energy + solvation free energy). The method was applied to adenine binding to small, synthetic receptors[14] and to cyclic ureas binding to HIV-1 protease.[15] Wang and coworkers[16] had earlier used a similar approach to compute relative binding free energies of different thrombin inhibitors. Kollman and coworkers[17–19] proposed a hybrid method (MM/PBSA) of computing absolute binding free energies. An explicit solvent simulation is used to generate many snapshots of the complex and, subsequently, the Poisson-Boltzmann equation and a surface-area equation are used to estimate the electrostatic and nonpolar contributions to binding, respectively. A number of other studies similar in spirit have been reported.[20–22]

In addition to the above fundamental approaches, more empirical computational methods have been proposed. Freire and coworkers[23] developed an empirical method for predicting the binding free energy from the polar and nonpolar surface area buried upon binding. Another empirical method includes electrostatic interactions calculated by a Molecular Mechanics force field.[24] Åqvist et al.[25,26] proposed the Linear Interaction Energy method, where the van der Waals and electrostatic interactions of the ligand with the protein in the complex and with water in its unbound state are computed by molecular dynamics (MD) simulations. The difference in these interactions between bound and solvated states is scaled by empirical factors and summed to give the binding free energy. The factor $\alpha = 0.5$ for the electrostatic interaction can be justified based on the linear response approximation of continuum electrostatics,[25] but the coefficient β for the van der Waals term seems to be entirely empirical. Studies of different systems showed that no set of coefficients is universally applicable. For example, it was shown that $\beta = 1$ works much better for avidin ligands,[27] whereas the value $\beta = 0.16$ was optimal in the original study.[25] Actually, a correlation has been found between the value of β and the hydrophobicity of the binding site.[28] Jorgensen and coworkers found that including the change in surface area and a different set of coefficients works better for nonnucleoside inhibitors of HIV reverse transcriptase[29] and FKBP12 inhibitors.[30] Sham et al.[31] used similar methods in the study of cyclic urea inhibitors of HIV protease.

The present work aims to analyze the contributions to the absolute binding free energy computed from first principles using MD simulations with an implicit solvation model.[32] First, a theoretical framework is given for the decomposition of the binding free energy into effective energy and entropic contributions. The effective energy changes upon binding are obtained by averaging the effective energy in the simulations. Particular emphasis is given to the protein and ligand reorganization energies. The pairwise decomposability of the energy function allows the determination of group contributions to the binding free energy. The translational and rotational entropies are obtained by computing the distributions of residual translational and rotational motions of the ligand in the protein and comparing with those in the standard state. The conformational entropy for the protein sidechains are obtained from probability distributions computed in MD simulations of the free and bound protein. The conformational entropy of the ligand is obtained by the number of freely rotatable bonds multiplied by values for the entropy per torsion suggested in the literature.

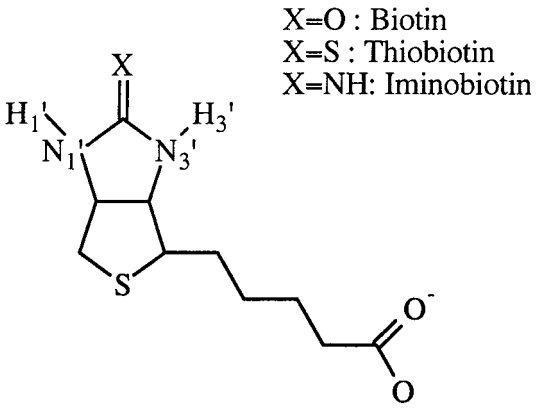

Fig. 1. Biotin and its analogs.

The system we chose for this study is biotin, biotin analogs (Fig. 1), and linear and cyclic peptides binding to streptavidin and avidin. This system was chosen because (a) with a dissociation constant of about 1 fM, it is among the strongest binding natural pairs known[33]; (b) it is widely used in technological applications, such as immunoassays, biosensors, or separations[34]; (c) crystal structures of this complex, of complexes of biotin with avidin and other proteins, and of other ligands with streptavidin, are available; (d) thermodynamic data for many of the above complexes are also available; and (e) theoretical studies with different methods have already been performed.

The biotin-streptavidin system has been the subject of a free energy simulation study[35,36] in which the biotin ligand was "annihilated" in the protein binding site and in the solvent. The calculated binding free energy was partitioned into contributions from van der Waals and electrostatic interactions. The packing effects in the protein binding site and the hydrophobic solvation effects for biotin in solution were both lumped into the van der Waals term. The contributions of individual residues were not analyzed, nor was the loss of translational, rotational, and conformational entropy upon binding considered.[37] A similar methodology with more state-of-the-art computational resources was recently applied to the same system by Dixit and Chipot.[38] Other studies by Dixon and Kollman[39] include a free energy simulation of streptavidin polar residue mutants, estimates of binding free energy of fluorinated biotin,[40] and a number of other ligands[18] by the MM/PBSA method, and application of the linear

interaction energy method to the binding of the same ligands.[27] Other groups[41,42] studied the kinetics of dissociation of biotin from streptavidin but did not attempt an analysis of binding thermodynamics.

THEORY

In this section we show how the binding free energy can be separated into energetic, solvation, and entropic contributions. The treatment is similar to that of folding thermodynamics of a single macromolecule.[32] An alternative statistical mechanical formalism is given by Gilson et al.[37] The main advantage of the present formalism is a more clear separation of the entropic contributions.

We start with a classical expression for the free energy of a system consisting of one macromolecule at infinite dilution in solution[32]:

$$A = A^{o} + kT \ln\left(\frac{\Lambda^{3M}}{V8\pi^2}\right) + \int p(\mathbf{q})W(\mathbf{q})d\mathbf{q} + kT\int p(\mathbf{q})\ln p(\mathbf{q})d\mathbf{q} \quad (1)$$

where A^{o} is the free energy of the pure solvent, and the second term is the ideal contribution from macromolecular translation and rotation. Λ is the thermal de Broglie wavelength, M is the number of macromolecular atoms, and V is the volume of the system (canonical ensemble). The third term in Eq. 1 is the average effective energy, which is equal to the average intramolecular energy plus the average solvation free energy ($W = E^{\text{intra}} + \Delta G^{\text{solv}}$), where $\mathbf{q}$ stands for the internal coordinates and $p(\mathbf{q})$ is the conformational probability distribution. The last term (divided by -T) is the configurational entropy of the macromolecule, including the conformational and vibrational entropy contributions.

If we have two macromolecules (A and B) instead of one, $\mathbf{q}$ is replaced by $\mathbf{q}^{A}$, $\mathbf{q}^{B}$, and $\mathbf{R}$, where $\mathbf{q}^{A}$ and $\mathbf{q}^{B}$ are the internal coordinates of A and B and $\mathbf{R}$ is a vector of six coordinates specifying the relative position and orientation of the two macromolecules:

$$A = A^{o}+kT\ln\left(\frac{\Lambda^{3M}}{V8\pi^2}\right) + \int p(\mathbf{q}^{A},\mathbf{q}^{B},\mathbf{R})W(\mathbf{q}^{A},\mathbf{q}^{B},\mathbf{R})d\mathbf{q}^{A}d\mathbf{q}^{B}d\mathbf{R} + kT\int p(\mathbf{q}^{A},\mathbf{q}^{B},\mathbf{R})\ln p(\mathbf{q}^{A},\mathbf{q}^{B},\mathbf{R})d\mathbf{q}^{A}d\mathbf{q}^{B}d\mathbf{R} \quad (2)$$

The probability distribution of the internal and relative external coordinates can be written as:

$$p(\mathbf{q}^{A},\mathbf{q}^{B},\mathbf{R}) = p(\mathbf{R})p(\mathbf{q}^{A},\mathbf{q}^{B}|\mathbf{R}) \quad (3)$$

where $p(\mathbf{q}^{A},\mathbf{q}^{B}|\mathbf{R})$ is the conditional probability distribution of finding the macromolecules at a certain configuration given their relative position and orientation. Now the free energy of the system can be written

$$A = A^{o}+kT\ln\left(\frac{\Lambda^{3M}}{V8\pi^2}\right) + \int p(\mathbf{R})d\mathbf{R}\int p(\mathbf{q}^{A},\mathbf{q}^{B}|\mathbf{R})W(\mathbf{q}^{A},\mathbf{q}^{B},\mathbf{R})d\mathbf{q}^{A}d\mathbf{q}^{B} + kT\int p(\mathbf{R})\ln p(\mathbf{R})d\mathbf{R} + kT\int p(\mathbf{R})d\mathbf{R}\int p(\mathbf{q}^{A},\mathbf{q}^{B}|\mathbf{R})\ln p(\mathbf{q}^{A},\mathbf{q}^{B}|\mathbf{R})d\mathbf{q}^{A}d\mathbf{q}^{B} \quad (4)$$

The third term is the average effective energy, the fourth term is the relative translational/rotational entropy, and the last term is the conformational entropy. The separation of the translational/rotational entropy from the conformational entropy in not unique because it depends on the choice for $\mathbf{R}$.[37] A reasonable choice for $\mathbf{R}$ would be the relative position of the two centers of mass ($\mathbf{r}$) and three Euler angles (ω) describing the orientation of the principal axes of one macromolecule with respect to the principal axes of the other.

The translational/rotational entropy term can be further split into translational and rotational terms if we assume that the translational and rotational degrees of freedom are decoupled, that is

$$p(\mathbf{R}) = p(\mathbf{r})p(\omega) \quad (5)$$

In that case, the fourth term of Eq. 4 becomes

$$kT\int p(\mathbf{r})\ln p(\mathbf{r})d\mathbf{r} + kT\int p(\omega)\ln p(\omega)d\omega \quad (6)$$

The integrals can be further split into a sum of terms for each of the three translational and rotational coordinates, if we assume that these coordinates are uncorrelated from each other.

The free energy of association of the two macromolecules is defined through the bimolecular equilibrium constant K:

$$\Delta G^{1M} = -kT\ln[K \cdot 1M], \quad K = \frac{\rho_{AB}}{\rho_A\rho_B} \quad (7)$$

where ρ_{AB}, ρ_A, and ρ_B are the molar concentrations (or activities for nonideal solutions) of the complex and the free molecules. The definition of "complex" (which configurations are considered "complex") in principle depends on the experimental technique used to measure K. In Eq. 7, the dependence of the free energy on the standard state has been explicitly indicated. The standard binding free energy is the free energy for the process of taking a molecule of A (or B), which is free to move in a volume V^{o} and placing it close to a molecule of B (or A), restricting its motion to configurations consistent with the definition of "complex." The standard state specifies the volume V^{o}. The common standard state of $1M$ corresponds to setting $V^{o} = V^{1M} = 1660$ Å^3 (1 L/mol = 1660 Å^3/molecule).

The binding free energy is the free energy difference between the bound state (AB) and the unbound state (A,B):

$$\Delta G^{1M} \approx \Delta A^{1M} = A(\mathrm{AB}) - A(\mathrm{A,Bin}V^{1M})$$
$$= -kT \ln \frac{p(\mathrm{AB})}{p(\mathrm{A,Bin}V^{1M})} = -kT \ln \frac{Z(\mathrm{AB})}{Z(\mathrm{A,Bin}V^{1M})} \quad (8)$$

where p( ) and Z( ) are the probability and the configurational partition function of a state, respectively. The free energies of the two states are given by Eq. 4 using the appropriate probability distribution, normalized for that state.

For the unbound state the expression can be simplified, because $\mathbf{q}^{\mathrm{A}}$ and $\mathbf{q}^{\mathrm{B}}$ are now decoupled ($p(\mathbf{q}^{\mathrm{A}},\mathbf{q}^{\mathrm{B}}|\mathbf{R}) = p(\mathbf{q}^{\mathrm{A}})$ $p(\mathbf{q}^{\mathrm{B}})$) and $p(\mathbf{R})$ is constant (from normalization $p(\mathbf{R}) = 1/V8\pi^2$). In addition, the two macromolecules do not interact, so that $W = W_{\mathrm{A}} + W_{\mathrm{B}}$. Therefore,

$$A(\mathrm{A,Bin}V) = A^{\circ} + kT \ln\left(\frac{\Lambda^{3M}}{V8\pi^2}\right) + kT\int p(\mathbf{q}^{\mathrm{A}})W_{\mathrm{A}}d\mathbf{q}^{\mathrm{A}}$$
$$+ kT\int p(\mathbf{q}^{\mathrm{B}})W_{\mathrm{B}}d\mathbf{q}^{\mathrm{B}} + kT\int \frac{1}{V8\pi^2}\ln\frac{1}{V8\pi^2}\,d\mathbf{R}$$
$$+ kT\int p(\mathbf{q}^{\mathrm{A}})\ln p(\mathbf{q}^{\mathrm{A}})d\mathbf{q}^{\mathrm{A}} + kT\int p(\mathbf{q}^{\mathrm{B}})\ln p(\mathbf{q}^{\mathrm{B}})d\mathbf{q}^{\mathrm{B}} \quad (9)$$

If the bound state is treated similarly, assuming that $p(\mathbf{R})$ is uniform within the allowed interval, the translational/rotational entropy contribution is

$$-kT \ln V_{\mathrm{AB}} \quad (10)$$

where V_{AB} is $\int_{\mathbf{complex}}\mathbf{dR}$ (the volume of configuration space corresponding to the complex). The change in translational/rotational entropy upon binding is:

$$-kT \ln (V_{\mathrm{AB}}/V_{\mathrm{A,B}}) \quad (11)$$

This is the approach taken by Finkelstein and Janin.[43]

Alternatively, the bound complex can be considered as virtually a single molecule. This is perhaps a more realistic view, because biomolecular complexes are very tight, exhibiting interfaces that are often indistinguishable from the interior of individual proteins.[44] In addition, the assumption of a uniform $p(\mathbf{R})$ is probably not valid. In this view, we can think of $\mathbf{R}$ as a set of six imaginary internal coordinates linking two arbitrary atoms, one from each macromolecule and avoid any separation between $\mathbf{R}$ and q^{A}, q^{B}. This is the approach taken by Tidor and Karplus,[45] who evaluated the translational/rotational entropy loss upon binding by performing vibrational analysis of the biomolecular complex. The advantage of this approach is that it uses quantum mechanical formulas to obtain thermodynamic properties (which is appropriate for the high frequency vibrations). The disadvantage is that it relies on the harmonic approximation.

METHODS

Approximations

In this work we start from Eq. 4 for the complex and Eq. 9 for the free state and make the following simplifying approximations. For the translational and rotational entropy, we assume that the six relative external coordinates $\mathbf{R}$ are uncorrelated from each other and compute probability distributions for each of them from MD simulations of the complex. For the conformational entropy in the complex, we assume that there is little variation in the conformational distribution with $\mathbf{R}$ ($p(\mathbf{q}^{\mathrm{A}},\mathbf{q}^{\mathrm{B}}|\mathbf{R}) \approx p(\mathbf{q}^{\mathrm{A}},\mathbf{q}^{\mathrm{B}})$. We further consider only dihedral angle distribution changes upon binding and assume that these distributions are uncoupled, so that the change in conformational entropy upon binding is the sum of the differences in the integrals $kT\int p(\mathbf{q}_{\mathrm{i}})\ln p(\mathbf{q}_{\mathrm{i}})d\mathbf{q}_{\mathrm{i}}$ between the bound and unbound state for each dihedral angle $\mathbf{q}_{\mathrm{i}}$ of the protein and the ligand. With these assumptions the conformational entropy in the complex (last term in Eq. 4) can be decomposed into contributions from A and B. The standard free energy change is then the difference between Eqs. 4 and 9:

$$\Delta G = \Delta W - T\Delta S^{\mathrm{trans/rot}} - T\Delta S^{\mathrm{confA}} - T\Delta S^{\mathrm{confB}} \quad (12)$$

where the terms on the right hand side are differences in the corresponding terms of Eqs. 4 and 9.

The change in effective energy (ΔW) can be broken down to finer contributions. Because our intramolecular energy function is pairwise additive and the solvation free energy a sum over group contributions, the effective energy of the system can be written:

$$W = E^{\mathrm{A}} + E^{\mathrm{B}} + E^{\mathrm{inte}} + \Delta G^{\mathrm{slvA}} + \Delta G^{\mathrm{slvB}} \quad (13)$$

The change in effective energy upon binding is

$$\Delta W = \Delta E^{\mathrm{A}} + \Delta E^{\mathrm{B}} + E^{\mathrm{inte}} + \Delta\Delta G^{\mathrm{slvA}} + \Delta\Delta G^{\mathrm{slvB}} \quad (14)$$

where ΔE^{A} and ΔE^{B} are the changes in the intramolecular energy of A and B, respectively, E^{inte} the interaction energy between A and B and $\Delta\Delta G^{\mathrm{slvA}}$ and $\Delta\Delta G^{\mathrm{slvB}}$ the change in solvation free energy of A and B upon binding.

Because the solvation free energy function is pairwise additive (see Eq. 16 below), the desolvation effect upon binding can be decomposed into the change in solvation of A due to changes in the positions of A atoms, the change in solvation of B due to changes in positions of B atoms, the change in solvation of A due to the approach of B atoms, and the change in solvation of B due to the approach of A atoms. Thus, an alternative way of writing ΔW is:

$$\Delta W = \Delta W^{\mathrm{A}} + \Delta W^{\mathrm{B}} + W^{\mathrm{inte}} \quad (15)$$

where ΔW^{A} includes ΔE^{A} and the change in solvation free energy of the groups of A due to other A atoms, and similarly for ΔW^{B}. W^{inte} includes E^{inte} and the change in solvation free energy of A and B due to intermolecular interactions.

Because of the inclusion of the solvation free energy changes, ΔW is not entirely enthalpic. The enthalpy of binding is obtained by subtracting the solvation free energy change and adding the solvation enthalpy change:

$$\Delta H = \Delta W - \Delta\Delta G^{\mathrm{slv}} - \Delta\Delta H^{\mathrm{slv}} \quad (16)$$

Energy Function

The change in effective energy in Eq. 12 is evaluated with the energy function EEF1, consisting of the CHARMM 19 polar hydrogen force field[46,47] augmented by an analytical model for the solvation free energy.[32] The solvation model assumes that the solvation free energy of a macromolecule is the sum of contributions from its constituent groups. The solvation free energy of each group is equal to the solvation free energy of that group in small model compounds[48] minus the amount of solvation it loses due to exclusion of solvent by other protein atoms around it.

$$\Delta G^{\mathrm{slv}} = \sum_{\mathrm{i}} \Delta G_{\mathrm{i}}^{\mathrm{slv}} = \sum_{\mathrm{i}} \Delta G_{\mathrm{i}}^{\mathrm{ref}} - \sum_{\mathrm{i}} \sum_{\mathrm{j} \neq 1} f_{\mathrm{i}}(r_{\mathrm{ij}}) V_{\mathrm{j}} \qquad (17)$$

The same model can be used to calculate solvation enthalpies, if the solvation free energy parameters are replaced by solvation enthalpy parameters.[32,49]

EEF1 has been tested extensively. It yields modest deviations from crystal structures upon MD simulations at room temperature, unfolding pathways in agreement with explicit solvent simulations, and it discriminates native conformations from misfolded decoys.[50] It has been used to determine the folding free energy landscape of a β-hairpin,[51] exploration of partially unfolded states of α-lactalbumin,[52] studies of protein unfolding,[53–56] identification of stable building blocks in proteins,[57] and analysis of the energy landscape of polyalanine.[58]

The partial charges for biotin, thiobiotin, and iminobiotin were adapted from the work of Miyamoto and Kollman.[35] To make the modeling of these ligands consistent with EEF1, we used united atoms for the aliphatic chain and a neutralized carboxylate group. The bonded and nonbonded parameters were obtained from similar atom types in Charmm19. The solvation parameters of the biotin atoms were obtained from similar atoms in EEF1 with one exception. The imino group in iminobiotin has a partial charge distribution more polarized than that of the closest EEF1 group (NC2). In this case, we determined a new solvation parameter by performing a free energy simulation (using the TSM module of CHARMM) changing the partial charges of C and N from 0.52 and −0.85 (corresponding to NC2) to 0.71 and −1.04. The calculated solvation free energy change was −15.7 kcal/mol and was added to the −10 kcal/mol solvation parameter for NC2.

Systems and Protocols

We studied the following systems: biotin-streptavidin, biotin-avidin, thiobiotin-avidin, iminobiotin-avidin, FSHPQNT (hereafter designated LIN)-streptavidin , c-CHPQGPPC (hereafter designated CYCL)-streptavidin. The starting points of the computations were the crystal structures with PDB codes: 1STP (biotin-streptavidin),[59] 1AVD (biotin-avidin), and 1AVE (unliganded avidin),[60] 1SLG (FSHPQNT-streptavidin), and 1SLE (c-CHPQGPPC-streptavidin).[61] Starting coordinates for unliganded streptavidin were obtained from 1STP after deleting the biotin coordinates. Only the residues visible in the crystal structures were included in the simulations. Residue numbering is based on the simulated residues (to obtain the standard numbering add 12 for streptavidin and 2 for avidin). Initial coordinates for thiobiotin and iminobiotin were generated from those of biotin. The tetramer coordinates were generated by using the CRYSTAL facility of CHARMM. The crystal operations for each space group were obtained from http://www-structure.llnl.gov/Xray/tutorial/spcgrps.htm. Each structure was first minimized with 300 ABNR steps and subjected to MD simulation for 1 ns (2-fs timestep). The ligands alone and the free protein tetramers were also simulated for 1 ns. All MD simulations were completely unrestrained.

Average energies were calculated from the saved coordinates over the last 900 ps of the simulations. Most difficult to calculate is the change in protein intramolecular energy, because it corresponds to the small difference between large numbers. We have tried the following three protocols for taking the difference in intramolecular energy of the protein between complex and unliganded state: (a) consider the whole protein, (b) consider only residues with atoms within 10 Å from any ligand atom, and (c) consider only residues with atoms within 5 Å from any ligand atom. As we progress from (a) to (c), the subtracted numbers and the statistical uncertainty become smaller, but we face the risk of neglecting possible contributions from residues farther than 5 or 10 Å. Calculations were performed independently for the four binding sites, and the results were averaged. The discrepancy among the four values gives an indication of the random error in the calculations. When average effective energies from different simulations are subtracted, a systematic error can arise if the average temperature in the simulations is not exactly the same. To correct for such small temperature differences, we consider different 2-ps blocks of the simulation and determine the slope of $\langle W \rangle$ versus $\langle T \rangle$ (a heat capacity-like quantity) and use it to bring the $\langle W \rangle$ to the same reference temperature (2–3 degrees at the most).

The biotin-streptavidin and CYCL-streptavidin complexes were also simulated in explicit solvent with stochastic boundary conditions. A water sphere of 36 Å was overlaid on one of the four binding sites, and all water molecules overlapping with protein atoms were deleted. All protein atoms beyond 36 Å from the chosen binding site were kept fixed. The biotin-streptavidin complex monomer was also simulated in periodic boundary conditions in a cube of 61.14 Å edge with either a 12 Å cutoff or Ewald summation.

Translational and Rotational Entropy ($\Delta S^{\mathrm{trans/rot}}$)

The center of mass coordinates of biotin was calculated and subtracted from the center of mass coordinates of the surrounding binding site. Using the center of mass of the binding site rather than that of the whole protein is, in principle, more appropriate because otherwise we would include any movements of the binding site with respect to the rest of the protein (which is essentially protein conformational/vibrational entropy) into the entropy of the ligand. If the center of mass of the whole protein is used, slightly smaller values for the translational entropy change

are obtained. The translational entropy change is calculated in two stages. First, the range of relative x, y, and z values defines a "box" of volume $\Delta x\Delta y\Delta z$ in which the ligand moves. Restricting the motion of the ligand in this box from the standard state of 1M costs

$$\Delta S_1 = R \ln \frac{\Delta x \Delta y \Delta z}{1660 \text{Å}^3} \quad (18)$$

An additional reduction in entropy results from the fact that the probability distribution within this box is not flat. From normalization ($\int_{\Delta x} p(x)dx = 1$), a flat distribution corresponds to $p(x) = 1/\Delta x$. Therefore,

$$\Delta S_2^x = -R\left\{\int_{\Delta x} p(x)\ln p(x)dx - \int_{\Delta x} (1/\Delta x)\ln (1/\Delta x)dx\right\}$$

$$= -R\left\{\int_{\Delta x} p(x)\ln p(x)dx + \ln \Delta x\right\} \quad (19)$$

and similarly for ΔS_2^y,ΔS_2^z. The translational entropy loss is then:

$$\Delta S^{\text{trans}} = \Delta S_1 + \Delta S_2^x + \Delta S_2^y + \Delta S_2^z \quad (20)$$

For the orientational entropy, we need to calculate the distribution of orientations of the ligand with respect to the binding site. To do that, we first use the CHARMM command ORIENT to align the binding site with the space-fixed frame of reference, and then we use ORIENT again to align the ligand. From the rotation matrix reported by CHARMM, we calculate the Euler angles.[62] The distribution of the three Euler angles is used to compute the entropy change as:

$$\Delta S^{\text{rot}} = -R\left\{\int_0^{180} p(\theta)\ln p(\theta)\sin\theta d\theta - \int_0^{180} p_\theta \ln p_\theta \sin\theta d\theta\right\}$$

$$-R\left\{\int_0^{360} p(\phi)\ln p(\phi)d\phi - \int_0^{360} p_\phi \ln p_\phi d\phi\right\}$$

$$-R\left\{\int_0^{360} p(\chi)\ln p(\chi)d\chi - \int_0^{360} p_\chi \ln p_\chi d\chi\right\} \quad (21)$$

where the constant values p_θ, p_ϕ, p_χ corresponding to free rotation are determined from normalization ($\int_0^{180} p_\theta \sin\theta d\theta = \int_0^{360} p_\phi d\phi = \int_0^{360} p_\chi d\chi = 1$).

Conformational Entropy (ΔS^{conf})

The conformational entropy change for protein side-chains upon ligand binding is estimated by comparing the distribution of χ angles in the complex and in the free protein for those side-chains that have atoms within 5 Å from any ligand atom (except alanines; rotations around methyl groups, which are represented as extended atoms in our force field, are unlikely to be affected). The integrals of $p\ln(p)$ of the two distributions are used to estimate the entropy change. This calculation assumes sufficient sampling of side-chain rotamers during the 1-ns MD simulations. The same approach is used for the backbone ϕ and ψ angles of the binding site loop (residues 34–40), which is thought to become ordered upon biotin binding.

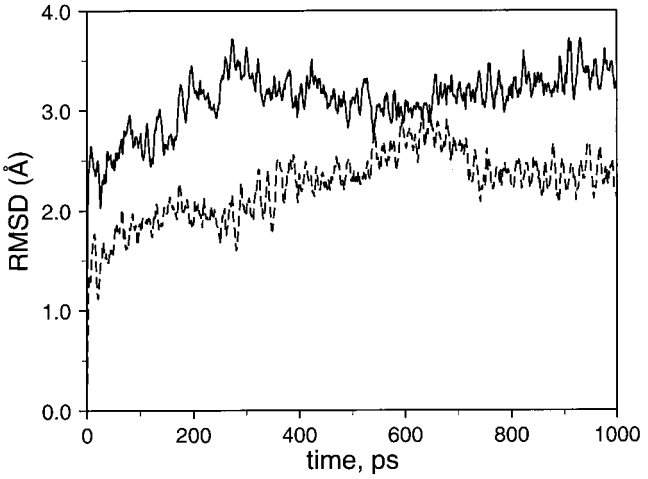

Fig. 2. Backbone RMS deviation from the crystal structure as a function of time for unbound streptavidin (solid line) and avidin (broken line) tetramers.

Because effective sampling of all the conformational space of a ligand is not possible in a short simulation, the conformational entropy of the ligand is based on the number of rotatable bonds it contains. The empirical scale of Picket and Sternberg[63] suggests 0.54 kcal/mol (TS at 300 K) for fixing a single bond. Melting of liquid alkanes suggests 0.45 kcal/mol for the same quantity.[64] Theoretical calculations gave 0.3 kcal/mol[14] and 0.48 kcal/mol[16] for the conformational entropy contribution per rotatable bond. Similar values have been obtained by Monte Carlo simulations of peptides[65] and energy maps of small organic molecules.[66] In this work, we adopt an average value 0.4 kcal/mol per rotatable bond.

RESULTS

Structural Stability

All systems studied in this work were subjected to fully unrestrained simulations at 300 K for 1 ns in their tetrameric form. We first compare the structure of the simulated systems to the crystal structures. The backbone RMSDs at the end of the simulations were modest: from 2.64 Å for biotin-streptavidin to 3.66 Å for biotin-avidin. Figure 2 shows the backbone RMSD versus time for the free streptavidin and avidin tetramers. These deviations are reasonable for systems of such size but give only a global measure of structural stability. By looking more closely at the binding site we observed significant shifts from the crystal structure.

In the biotin-streptavidin crystal structure (1STP), the ureido O of biotin hydrogen bonds to Asn 11, Ser 15, and Tyr 31 side-chains, the ureido N1′ hydrogen bonds to Asp 116 and the N3′ to Ser 33 [Fig. 3(a)]. At the end of the dynamics, we observe a multitude of binding modes of

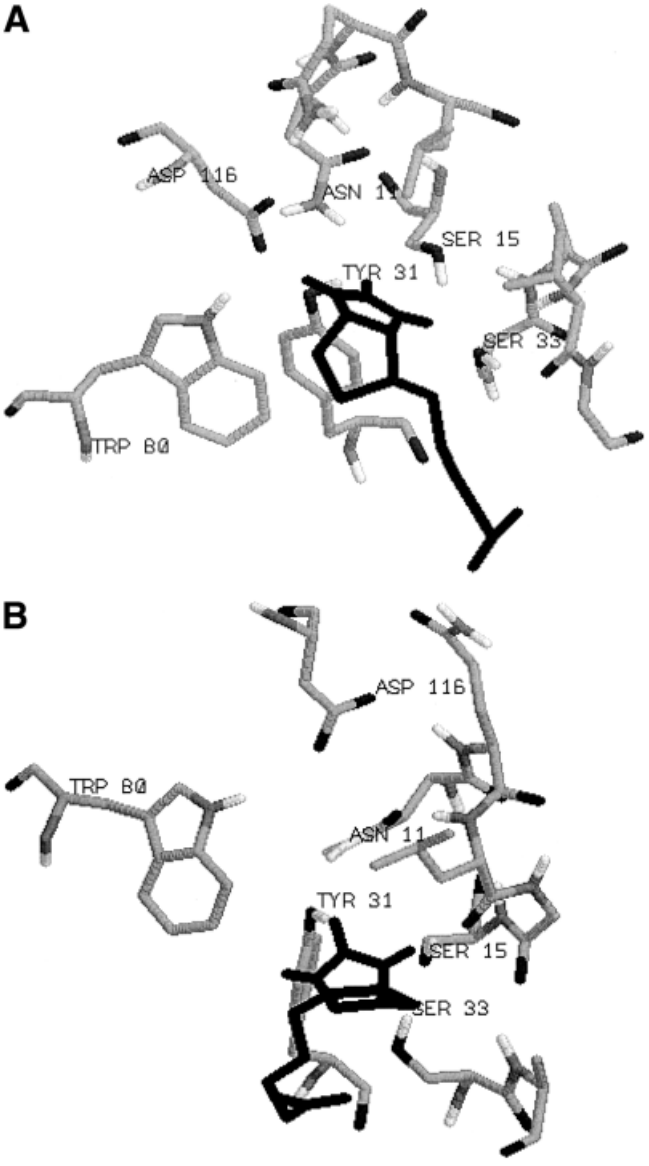

Fig. 3. Biotin in the binding site of streptavidin (**a**) crystal structure (**b**) after 1 ns of MD simulation

biotin in the four binding sites [e.g., Fig. 3(b)]. In all four sites, the hydrogen bond to Asp 116 breaks off quite early, and Asp 116 then hydrogen bonds to Asn 11 or Gln 12 or Trp 80 HE. The ureido O of biotin hydrogen bonds to Ala 34 backbone NH, or to Asn 11 and Tyr 31, or to Ser 15 HG, or to Trp 80 HE. The N1′ hydrogen bonds to Ser 15 and Tyr 31 or forms no hydrogen bonds. The N3′ hydrogen bonds to Val 35 backbone CO, or Leu 13 CO, or Ser 33 OG. These seem to be plausible interactions, and it is difficult to see why the crystal structure should be the only binding mode.

To test whether these structural shifts are an artifact of the implicit solvation model, we performed explicit solvent simulations by using the CHARMM22 all-hydrogen force field. A 1-ns simulation was performed with stochastic boundary conditions[67] with a sphere of 36 Å radius centered on one binding site. All protein atoms beyond 36 Å from the chosen binding site were fixed. We found that the Asp116-biotin hydrogen bond also breaks in this simulation, and the Asp116 moves toward the solvent. The ureido O maintains its hydrogen bond to Tyr 31 but breaks the one with Asn 11 and replaces the one with Ser 15 by one with Ser 33. The ureido N3′ replaces the hydrogen bond with Ser 33 by one with the backbone CO of Val 35. Two other simulations studied the streptavidin-biotin monomer in a box of 61.14 Å edge with periodic boundary conditions. One used the minimum-image convention and a cutoff of 12 Å. In this simulation (250 ps), the Asp 116 hydrogen bond to biotin remained, but the hydrogen bonds of the ureido O with Tyr 31 and Ser 15 broke. The second simulation used Ewald summation. After 350 ps in this simulation, the ureido O distance from Asp 116 and Ser 15 hydrogens is 2.5 and 2.95, respectively. Ser 33 moves toward the solvent and N3′ hydrogen bonds to Val 35 backbone CO. After an additional 100 ps, Asp 116 completely broke the hydrogen bond with biotin and moved toward the solvent. We also used the program TINKER (http://dasher.wustl.edu/tinker) to perform a simulation of the same system using the AMBER force field and a 9 Å cutoff in a 49.264 Å cubic box. After 150 ps, the hydrogen bond of N1′ to Asp116 was maintained, but the hydrogen bonds to Tyr 31, Ser 15, and Ser 33 were broken, as above for the CHARMM simulation. Thus, even in explicit solvent simulations, we observe significant shifts from the crystal structure. It is noted that Dixon and Kollman[39] reported that harmonic restraints on the backbone atoms were necessary to maintain apo-streptavidin close to the crystal structure.

In contrast, the binding mode of biotin to avidin is well maintained in the simulations. In avidin, Asp 116 is replaced by Asn, Ser 33 by Thr, Gly 36 by Thr, Asn 37 by Ala, Ala 38 by Thr, Ala74 by Phe, Glu12 by Asp, and Trp 80 by Phe. In the last replacement, the fact that Phe cannot serve as an alternative hydrogen-bonding partner to biotin may help to maintain the crystallographic hydrogen bonds. To assess the effect of Asp replacement, we performed a simulation of a mutant streptavidin (Asp116→Asn and Glu12→Asp, with which the Asp 116 was occasionally interacting). The breaking of the hydrogen bond with 116 also occurred in this mutant.

Some shifts are observed in the thiobiotin and iminobiotin simulations. In thiobiotin, the hydrogen bond between N1′ and Asn 116 breaks in all four binding sites, and N1′ forms other hydrogen bonds (e.g., to Tyr 31 or Ser 15 side-chains). The N3′ hydrogen bonds to either Tyr 31 OH, or Val 35 CO or Val 36 CO. In iminobiotin, the hydrogen bond between N1′ and Asn 116 breaks in one of the four sites, and N1′ hydrogen bonds to Asn 11 instead. N3′ hydrogen bonds to Tyr 31, or Thr 33 OG, or Val 35 CO. The imino group hydrogen bonds to Ser 15 HG or Asn 10 or Tyr 31.

Significant shifts are also observed in the peptide ligand simulations. In the crystal structure of LIN-streptavidin, the following hydrogen bonds are observed: peptide Asn5 with Asp116 and Asn11 HD, peptide Gln4 with Thr 87 HG, and peptide His 2 with Ser 76 HG [Fig. 4(a)]. After the

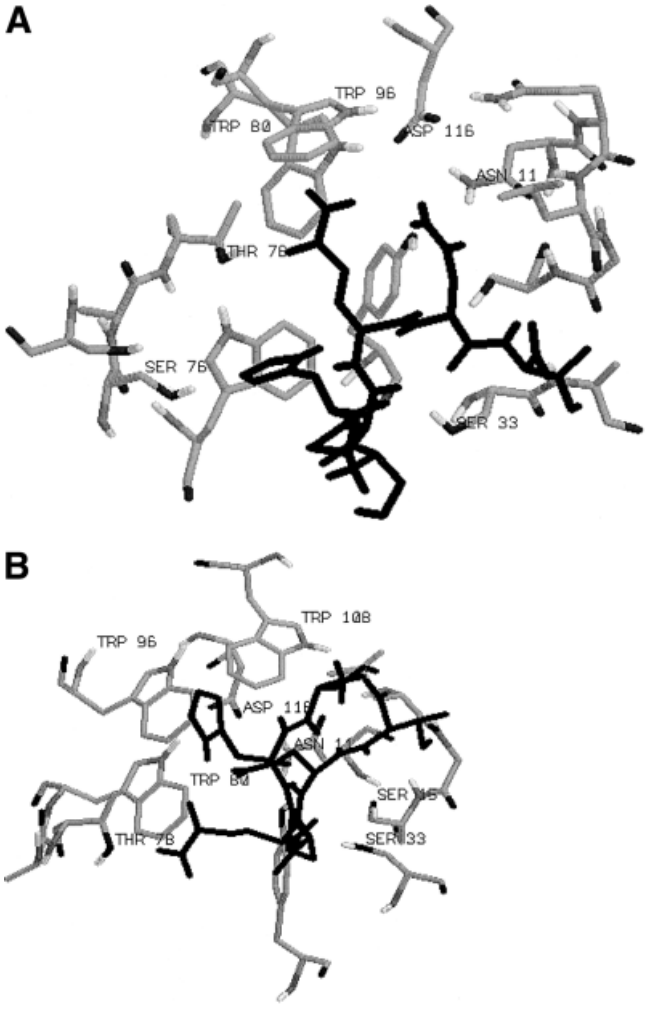

Fig. 4. The LIN peptide (FSHPQNT) in the binding site of streptavidin (**a**) crystal structure (**b**) after 1 ns of MD simulation

simulation, we observe a multitude of other interactions [e.g., Fig. 4(b)]. The hydrogen bond between peptide His 2 and Ser 76 is broken in all cases. The hydrogen bond between peptide Asn 5 and Asp 116 is maintained in two sites and broken in the other two sites. It is noted that for LIN-streptavidin, different structures were reported by two different crystallographic studies,[61,68] but the structural shifts we see are larger than the crystallographic differences. Similar observations are made in the CYCL-streptavidin simulation. A 1-ns explicit solvent simulation of CYCL was also performed with stochastic boundary conditions. We found some deviations in this simulation too. The hydrogen bond between peptide His 2 and Ser 76 OG was broken. In comparing explicit and implicit solvent simulations, note that conformational changes are slower in explicit solvent because of the friction and caging effects of the solvent.[53]

Finally, it is noted that the above structural shifts depend on protein flexibility. When we repeated the simulation of LIN keeping the protein fixed, the ligand stayed close to the crystal structure in all binding sites.

Protein-Ligand Interaction Energy

The protein-ligand effective interaction (W^{inte}, 2nd column of Table I) varies from -29 kcal/mol for biotin-streptavidin to -70 kcal/mol for LIN-streptavidin. The larger interaction energy for LIN versus biotin is largely a result of ligand size: LIN has 74 (extended) atoms and biotin 18. This quantity includes favorable contributions from van der Waals and electrostatic interactions and an unfavorable contribution from the mutual desolvation effect (i.e., the amount of solvation that the protein loses because of the presence of the ligand and vice versa). For example, the -36 kcal/mol for biotin-avidin include -29 kcal/mol from van der Waals, -29 kcal/mol from electrostatics, and $+22$ kcal/mol from desolvation. The desolvation is positive because it is dominated by the desolvation of polar groups (the desolvation of nonpolar groups is favorable but smaller in magnitude). The electrostatic energy can be fully assigned to the polar groups. Its magnitude approximately cancels the contribution from desolvation. About half of the van der Waals energy arises from nonpolar-nonpolar interactions, and the rest from nonpolar-polar interactions. Therefore, it is clear that the bulk of W^{inte} is contributed by the nonpolar groups, in agreement with previous studies.[35] A similar conclusion was reached in an analysis of protein folding stability.[69]

The group additivity of EEF1 allows us to decompose W^{inte} into residue contributions. For example, the -50 kcal/mol of CYCL in one binding site can be broken up into contributions of each residue of the peptide: ACE: -6.6, CYS: -0.6, HIS: -9.4, PRO: -9.7, GLN: -19.7, GLY: -5.5, PRO: -2.6, PRO: -1.6, CYS: -2.0, CBX: -0.15 kcal/mol.

An analogous decomposition of W^{inte} can be made in terms of the protein residues. For example, the larger W^{inte} of biotin complexed to avidin versus streptavidin can be analyzed in terms of residue contributions. We do this on minimized structures, rather than average over the dynamics trajectory, because of the observed deviation from the crystal structure in biotin-streptavidin (see above). The residue contributions are given in Table II. The total W^{inte} of biotin with the protein is -54.8 kcal/mol for avidin (average over four sites: -54.5) and -50 kcal/mol for streptavidin (average over four sites: -50.6). This difference is similar to the difference between the dynamics averages. The corresponding values for iminobiotin and thiobiotin in avidin (averages over the four sites) are -51.9 and -47 kcal/mol, respectively.

These results suggest that the more favorable interaction of biotin with avidin can be attributed to a few key residue substitutions: (a) replacing Asp116 by Asn increases the interaction energy because the desolvation cost of Asp is higher than that of Asn, (b) replacement of Gly 36 by Thr 36, (d) replacement of Ala 38 by Thr 38 (which interacts with the biotin tail COO^-), (e) replacement of Tyr 71 and Ala 74 by Phe 70 and Ser 71, and (f) replacement of Trp 80 by Phe 77 . It would be interesting to perform site-directed mutagenesis of the above residues in streptavidin to see if enhanced affinity is achieved.

TABLE I. Contributions to the Binding Free Energy (kcal/mol, T = 300 K)

Complex	$W^{inte\ a}$	ΔW^{lig}	ΔW^{prot} bs5	ΔW^{prot} bs10	ΔW^{prot} all	$-T\Delta S^{trans}$	$-T\Delta S^{rot}$	$-T\Delta S^{side}$	$-T\Delta S^{loop}$	$T\Delta S^{conf,lig}$	ΔG^{calc}	ΔG^{exp}
Strep-Biotin	−29	0	15	18	10	2.2	2.3	0.3	−1.9	4 * 0.4	−15 to −7	−18.3
Avid-Biotin	−36	0	9	11	21	2.8	2.7	1.7	−0.6	4 * 0.4	−19 to −7	−20.4
Avid-SBiotin	−31	0.5	9	15	20	2.5	2.0	−1.2	−1.1	4 * 0.4	−18 to −7	−16.9
Avid-IBiotin	−34	1	10	16	15	3.0	1.9	0.1	−0.5	5 * 0.4	−17 to −11	−14.3
Strep-LIN	−70	11	28	36	29	2.6	3.2	2.8	−1.6	27 * 0.4	−13 to −5	−5.32
Strep-CYCL	−50	7	16	17	11	2.3	2.1	3.4	−0.7	7 * 0.4	−22 to −16	−8.4
Error bars:	±2–6	±0.2–2	±3–8	±7–15	±15–20	±0.2–1	±0.2–1	±2–3	±2–3			

[a]The columns are, in sequence: ligand-protein effective interaction, ligand reorganization effective energy, protein reorganization effective energy (all: whole protein, bs5: binding site within 5 Å from ligand, bs10: binding site within 10 Å from ligand), translational entropy, rotational entropy, loop backbone entropy, conformational entropy of ligand, total estimated free energy, experimental binding free energy.[61,68,89,90]

TABLE II. Effective Interactions of Streptavidin and Avidin Residues With Biotin[†]

Avidin		Streptavidin	
Asn 10	−0.4	Asn 11	−5.0
Leu 12	−0.9	Leu 13	−0.8
Ser 14	−7.5	Ser 15	−6.9
Tyr 31	−7.9	Tyr 31	−7.7
Thr 33	−3.6	Ser 33	−3.9
Ala 34	−0.3	Ala 34	−0.4
Val 35	−1.2	Val 35	−1.0
Thr 36	−2.4	Gly 36	−1.6
Ala 37	−3.8	Asn 37	−3.9
Thr 38	−5.5	Ala 38	−0.6
Trp 68	−2.6	Trp 67	−3.3
Phe 70	−1.2	Tyr 71	0
Ser 71	−4.4	Ala 74	−0.1
Ser 73	−1.4	Ser 76	−4.4
Thr 75	−0.2	Thr 78	−0.1
Phe 77	−1.1	Trp 80	−0.6
Trp 95	−1.4	Trp 96	−1.8
Leu 97	−2.0	Leu 98	−2.3
Asn 116	−5.9	Asp 116	−4.2
Trp 108	−2.5	Trp 108	−2.6
Sum	−55.6		−51.3

[†]Residue numbering is based on the simulated residues (to obtain the standard numbering, add 12 for streptavidin and 2 for avidin).

Ligand Reorganization Energy

To calculate the ligand reorganization (effective) energy, we performed simulations of the ligands alone and compared their energies to the intramolecular energy of the bound ligands. This calculation assumes that the sampling of conformational space by regular MD is sufficient for the convergence of $\langle W \rangle$. To check this assumption, we repeated the simulations of the ligands starting with different random numbers for the assignment of initial velocities and, for the linear peptide, from a fully extended conformation. In all cases, the average $\langle W \rangle$ values were within 1 kcal/mol from each other.

For biotin, the ligand reorganization energy is essentially zero. Biotin is a small molecule that cannot fold onto itself and can bind to (strept)avidin without sustaining any significant intramolecular strain. The ligand reorganization energy is slightly positive for thiobiotin and iminobiotin. For the peptides, the ligand reorganization energy is substantial. This is so because when unbound the peptides can relax their structures to maximize the interactions between their own atoms. This effect is larger for the more flexible linear peptide. When they are bound to the protein, maximizing the interactions with the protein atoms leads to a compromise in the intramolecular energies.

Protein Reorganization Energy

The protein reorganization energy is the most difficult to calculate because it corresponds to a small difference between large numbers. We followed three approaches. The straightforward approach is to first calculate ΔW as the difference between W(complex) and W(apoprotein) + W(isolated ligand) and then obtain ΔW^{prot} from

$$\Delta W = W^{inte} + \Delta W^{prot} + \Delta W^{lig} \Rightarrow \Delta W^{prot} = \Delta W - W^{inte} - \Delta W^{lig} \quad (22)$$

This is the value given under the column "all" in Table I. An alternative approach was to focus on the binding site and calculate the intramolecular energy of the binding site residues in the complex and in the apoprotein. We used two definitions of the binding site: (a) the residues with atoms within 5 Å from the ligand (under column bs5) and (b) the residues with atoms within 10 Å from the ligand (under column bs10).

There is considerable variability in the values obtained by the three methods and from the four binding sites (this gives rise to the large error bars). As expected, the statistical uncertainty decreases with the portion of the protein considered but is still substantial even for the 5 Å binding site definition. What is consistent, however, is that the protein reorganization energy is large and positive. The smallest value obtained is 9 kcal/mol and the largest 36 kcal/mol for the LIN peptide. At first, these values for ΔW^{prot} seem to be large, especially for a binding side that does not undergo large conformational changes upon ligand binding (other than the closing loop). However, such values can easily arise by adding a few hydrophobic interactions and hydrogen bonds. For example, Ser 15 in streptavidin hydrogen bonds to biotin when it is in the binding site, but in its absence, it reorients to hydrogen bond to the Ala 34 backbone amide. As a result, the average effective interaction between Ser 15 and Ala 34 is −0.2 kcal/mol in the complex and −3.1 kcal/mol in the

apoprotein. The largest contributions to ΔW^{prot} come from van der Waals, electrostatic, and desolvation interactions. For example, the 15 kcal/mol for streptavidin/biotin (bs5) is decomposed into +10.9 kcal/mol van der Waals, + 14.1 kcal/mol electrostatics, −8.5 kcal/mol desolvation, and −1.5 from bonded terms.

To confirm that these protein reorganization energies are not influenced by any artifactual shifts from the crystal structure upon MD simulation, we repeated the calculation for biotin-avidin by using the energy-minimized crystal structures of the complex (1AVD) and the apoprotein (1AVE). The average intrabinding site effective energy (5 Å definition) in the four binding sites was −46.6 kcal/mol in the apoprotein and −34.2 kcal/mol in the complex, giving a ΔW^{prot} of 12.4 kcal/mol. This is similar to the values obtained by averaging over the MD trajectories.

Translational and Rotational Entropy

The translational (1*M* standard state) and rotational entropy are calculated to be quite small, between 2 and 3 kcal/mol each. As expected, they are negative, and thus oppose binding. For the translational and rotational entropy, Finkelstein and Janin estimated 6.9 kcal/mol and 7.2 to 9 kcal/mol, respectively. These values are larger than those calculated here because the amplitudes of motion assumed by these authors are smaller than those found in the present simulations. They assumed a translational amplitude of 0.25 Å, whereas we find typical translational amplitudes of 4–6 Å. They also assume a rotational amplitude of 2–5 degrees, whereas we find much broader distribution of the Euler angles. Our estimates for $\Delta G^{trans/rot}$ are somewhat smaller than the 7 kcal/mol calculated by Luo and Gilson[14] for the binding of adenine to synthetic adenine receptors and by Hermans and Wang[8] for the binding of benzene to a lysozyme cavity. Perhaps this is due to a more flexible binding site or to implicit solvation, which allows a more extensive sampling of conformational space for the same amount of simulation time. For biotin-streptavidin, Dixit and Chipot[38] estimated a translational entropy loss of about 4.5 kcal/mol based on the effective volume sampled during an MD simulation and 1.8 kcal/mol from Free Energy Perturbation or Thermodynamic Integration with restraints. Baginski et al.[21] obtained a $\Delta G^{trans/rot}$ between 4 and 9 kcal/mol from free energy calculations and MD simulations, respectively. Our estimates are close to the value 6.2 kcal/mol predicted by fitting protein complexation free energies to a surface area-based model.[70] That the value for the translational entropy calculated here is similar to the so called "cratic" entropy term (Rlnx = Rln(1M/55.5M) = − 8 e.u., or 2.4 kcal/mol at 300 K)[71] is a coincidence. The translational entropy should depend on the concentration of the solute but not on the density of the solvent.[37]

Ligand Conformational Entropy

The ligand conformational entropy is not calculated directly but estimated on the basis of the number of torsional degrees of freedom and the value 0.4 kcal/mol per torsion. As expected, it is largest for the linear peptide and smallest for biotin. This is a rough calculation and should be improved in the future by performing systematic explorations of conformational space, whenever possible.[14]

Protein Side-Chain Conformational Entropy

The contribution to the binding entropy from the conformational freedom of protein side-chains was estimated by calculating side-chain dihedral angle distributions in the complex and in the free protein. Only those side-chains with atoms within 5 Å from the ligand in the crystal structure were considered. This results to 15 side-chains for biotin/streptavidin (30 dihedrals), 18 for biotin/avidin (36 dihedrals), and 20 for linear and cyclic peptides (44 dihedrals). The χ_1 and χ_2 distributions were assumed independent from each other. The total entropy was obtained by summing the contributions of each dihedral angle.

The results are shown in Table I under ΔS^{side}. This entropic contribution is rather small, in agreement with earlier calculations.[11] The entropy change is negative, except for thiobiotin, where it is small and positive. This means that in these two cases the protein side-chains have more conformational freedom in the complex than in the apoprotein. This finding would have been surprising a few years ago. However, NMR experiments in some systems have shown increased disorder of the backbone[72] or the side-chains[73] upon ligand binding, although in the latter study most of the increase is observed away from the binding interface.

Looking at the contributions of individual angles we see that some side-chains make a positive contribution to the entropy. For example, in the biotin-avidin complex, of the 36 side-chain dihedral angles considered, 25 exhibit negative entropy change and the rest are positive. Among the side-chains with a positive entropy change upon binding are the two dihedral angles of Asn 116 and the two of Trp 108. However, the side-chain entropy results suffer from a large statistical uncertainty (there is a large variation of the results in the four binding sites due to limited sampling), and firm conclusions can only be reached by more lengthy simulations or systematic sampling of side-chain rotamers.

Loop Backbone Entropy

The (strept)avidin binding site has a loop that appears to be disordered in the apoprotein and become ordered upon biotin binding.[59,74] This order-disorder transition should contribute negatively to the entropy of binding and unfavorably to the binding free energy. In this work, we attempted to estimate the contribution of the loop to the entropy of binding following the same methodology as for the protein side-chains above. We calculated the distributions of the backbone ϕ and ψ angles in the apoprotein and the complex for seven loop residues (from residues 34–40). Each angle is assumed independent of the others, and the total entropy contribution is the sum of those calculated for each angle. The results are shown in Table I (ΔS^{loop}). It is surprising that the calculated entropy contribution upon binding is positive (i.e., the fluctuations of the backbone

dihedrals are larger in the complex than in the apoprotein).

This result should be interpreted with caution, because the 1-ns MD simulations may not provide enough time for the loop to sample greatly different conformations. We also performed a simulation starting from another crystal structure of apo-streptavidin (1SWC)[74] where in two of the four monomers the loop is visible in an "open" conformation. The ϕ and ψ angles in these loops have quite different values from those in 1STP, where the loop is closed. MD simulations of this structure showed that the loops remained in the open conformation and the dihedrals did not change substantially. Apparently, there is a significant barrier to interconversion between the two loop conformations that cannot be traversed in the 1-ns simulations. This interconversion should be further studied in the future by using reaction path-mapping methods.

Binding Free Energy and Enthalpy

Obviously, the large uncertainty in the protein reorganization energy allows us only to estimate a range for the binding free energy. Nevertheless, the resulting free energies are of the correct order of magnitude. In most cases, the predicted range brackets the experimental values. The binding of biotin to streptavidin tends to be somewhat underestimated. The largest discrepancy is observed for CYCL binding to streptavidin, where the predicted range is significantly more negative than the experimental value. Possible origins of the discrepancies are considered in the Discussion.

Binding enthalpies can be obtained via Eq. 16. The $\Delta\Delta G^{\mathrm{slv}}$ and $\Delta\Delta H^{\mathrm{slv}}$ are obtained by subtracting the solvation free energy or enthalpy of the free protein and ligand from that of the complex (whole protein method). The results are of similar agreement to experiment as the free energy results. For example, for the LIN complex, $\Delta\Delta G^{\mathrm{slv}} = +35$ kcal/mol and $\Delta\Delta H^{\mathrm{slv}} = +49$ kcal/mol. This gives a ΔH in the range of -20 to -28 kcal/mol. The experimental value is -19.34 kcal/mol.[68] For biotin-avidin, $\Delta\Delta^{\mathrm{slv}} = +14$ kcal/mol and $\Delta\Delta H^{\mathrm{slv}} = +19$ kcal/mol. This gives ΔH in the range of -10 to -22, whereas the experimental value is -21.5 kcal/mol.[33]

DISCUSSION

Prediction of binding affinities is a long-standing goal from both a fundamental and a practical viewpoint. In the field of rational drug design, the ability to predict the binding affinity of a candidate molecule without having to synthesize it would save a lot of time and resources. Although the theory of binding seems to be well understood (see Theory section and Ref 37), implementation of these ideas for practical calculations is difficult. Accurate predictions require accurate force fields, accurate solvation free energy estimates, and realistic sampling of conformational space.

In this work, we propose a direct methodology of obtaining estimates of absolute binding free energies and its decomposition into energetic and entropic components. At this point, the method is most useful for obtaining physical insights, rather than accurate values. The advantages of the present method are that it is applicable to systems of any size and that it provides a direct separation of the components of the binding free energy. The lack of precision is probably not inherent in the method but in the complexity of the systems studied. Other methods, such as free energy perturbation, applied to the same systems would probably face similar precision problems. One disadvantage of the method is that it relies on regular dynamics simulations to sample conformational space, which may not be sufficient in case of slow rearrangements. For example, the conformational entropy of a flexible ligand cannot be calculated by such regular simulations. The precision could in principle be improved by performing longer simulations.

Our major finding is that the protein reorganization energy makes a large and positive contribution to the binding free energy, even for a relatively well preorganized binding site. This results from the fact that binding to the ligand causes some compromise in intraprotein interactions. Even though these energy losses may be small, they are significant compared to the total binding free energy, which is usually <20 kcal/mol. Most work on binding affinities so far has either assumed a rigid protein or neglected the contribution of protein reorganization. Honig and coworkers[11,12] recognized the potential importance of this quantity but did not calculate it. Horton and Lewis[70] attributed the difference in binding energy of BPTI to trypsin versus trypsinogen (about 10 kcal/mol) to the energy required to order the active site residues in trypsinogen. Hermans's method[8] includes implicitly the contribution of protein reorganization but will not be easy to apply to more complex, flexible solutes. Gilson and coworkers[14] have not so far considered protein flexibility. In Kollman's work,[18] identical configurations are used for the bound and the unbound protein; therefore, the protein reorganization energy is zero. What compensates for the lack of this contribution in their calculations is a large solute entropy contribution (see below). However, in a study of protein-RNA binding, where the structures of both the RNA-bound and free protein are known, the MM/PBSA method was used to calculate a 10 kcal/mol reorganization energy for the protein.[19] This value is similar to those obtained here, although the protein undergoes a more substantial conformational change in that system. Jayaram et al.[22] applied a similar method to calculate the energy of reorganization of DNA upon binding a protein and obtained the value 63 kcal/mol. Baginski et al.[21] obtained a DNA reorganization energy of $+32.2$ kcal/mol upon binding of an intercalating antibiotic. Noskov and Lim[20] found that protein reorganization is important but did not explicitly calculate protein reorganization energies. Our finding is that a significant protein reorganization energy can arise even from slight adaptation of the protein sidechains, without any visible conformational change. It would be worth calculating this quantity with other force fields and implicit solvation models to obtain a "second opinion" on its magnitude.

The protein reorganization energy tends to increase with ligand size. This scaling compensates for the scaling

of the favorable W^{inte} with ligand size. Another unfavorable contribution that scales with ligand size is the ligand conformational entropy and ligand reorganization energy. These compensating favorable (W^{inte}) and unfavorable (ΔW^{prot}, ΔW^{lig}, $\Delta S^{conf,lig}$) contributions combine into a ΔG that shows little correlation with ligand size.

The ligand reorganization energy can also be significant for flexible ligands. This quantity may exhibit some compensation with the ligand conformational entropy loss. The more structured a ligand is, the higher its reorganization energy will be and the smaller the conformational entropy cost will be. Ligand reorganization energies were recently calculated by using a Molecular Mechanics force field and the Generalized Born/Surface Area model.[75] For small ligands, the reorganization energy (excluding solvation) was <3 kcal/mol. For larger ligands, the authors felt that the energy function and/or the experimental structures are not reliable enough. The ligand reorganization energy was also considered by Vajda et al.,[24] but the reported results include the transfer free energy of the ligand from water to octanol and cannot be directly compared to ours. Vieth et al.[76] compared receptor-bound structures of small ligands to energy minima resulting from systematic search of conformational space in solution, but their emphasis was on structure rather than energetics.

The translational and rotational entropy losses were found to be smaller than suggested in the past because of substantial residual movements of the bound ligand. These values are significantly lower than estimates based on normal mode analysis of the solute.[18,20,45] The values obtained by normal mode analysis may be too large because of the harmonic approximation or because the analysis is done in vacuum, which may underestimate the range of movements available for the ligand in the complex. The protein side-chain and backbone entropy contribution is small and can sometimes increase upon ligand binding, a rather counterintuitive result. The fact that $\Delta S^{trans/rot}$ does not vary much for different ligands suggests that its calculation could be avoided in approximate studies, and a standard value could be used instead.

When the individual components are summed, the resulting free energies are of the correct order of magnitude. This suggests that our understanding of binding thermodynamics and the energy models we use are qualitatively correct. The level of (in)accuracy here is similar to other studies of absolute binding affinities in biological systems.[19,22] The most significant systematic discrepancy is for CYCL, whose affinity is overestimated. A likely origin of this is the energy function. The hydrogen bonds and interactions between ionizable residues in EEF1 may be somewhat too strong. The distance-dependent dielectric seems to overestimate the strength of the interactions near the protein surface and underestimate it in the protein interior.[77] Some weakening of solvent-exposed hydrogen bonds can easily bring the calculated free energy closer to experimental results. This would not affect so much the binding free energy of biotin because it forms few hydrogen bonds, and they are buried deeply in the binding site.

Another theoretically possible reason that ΔG for CYCL and LIN are overestimated could be that these peptides interact with each other in the unbound state. If such interactions exist, then ΔW^{lig} should be more positive because, to bind the protein, the ligands would have to break their interactions with other ligand molecules. Ligands with greater tendency to self-associate would exhibit lower affinity for a protein. This possibility could be excluded if careful experiments showed lack of ligand self-association under the conditions of the experimental binding affinity measurements. In addition, in these approximate calculations, we neglect protein entropy contributions beyond the residues in contact with the ligand, although NMR experiments showed changes in mobility upon ligand binding throughout the protein.[72] If such changes take place, they could have a substantial effect on the binding thermodynamics.

A qualitative conclusion concerning the biotin-(strept)-avidin complex is that the exceptionally high affinity is primarily due to the strong interaction energy (W^{inte}) between ligand and protein considering the small size of biotin. The interaction energy per atom for biotin binding to streptavidin and avidin is -1.6 kcal/mol and -2 kcal/mol, respectively, whereas for the LIN peptide is -0.94 kcal/mol. A large part of this interaction energy comes from van der Waals interactions. In other words, the strong affinity is due to the snug fit between ligand and protein, in agreement with earlier studies.[35,36] Another factor in the strong binding is the negligible ligand reorganization energy of biotin. The conformational energy of biotin in the binding site is similar to that in solution. The small ligand size also makes the ligand conformational entropy loss upon binding small. Some additional stabilization in streptavidin may arise from polarization of the biotin ureido group from the charged Asp116, which is not included in classical force fields.[59] Quantum mechanical calculations (not reported here; see http://blakey.sci.ccny.cuny.edu /biotin) suggest that the polarization of biotin by Asp 116 may contribute a couple of kcal/mol extra to the binding free energy in streptavidin.

A qualitative understanding of the fundamentals of binding may be useful in developing less rigorous but "pragmatic" methods to estimate binding affinity. One interesting observation in this regard is that relative binding free energies between similar ligands could be obtained by energy minimizations, rather than costly MD simulations. The validity of such an approach will be explored in future studies. Another practical method is the "linear interaction energy" approach.[25,26] The difficulty with this approach seems to be that the empirical parameters α and β are not transferable from one system to another. It is likely that scaling the β parameter corrects for the omission of other contributions to the binding free energy. A smaller β parameter could compensate for a larger reorganization energy cost. Calculations of ligand and protein reorganization energies in different systems

could perhaps explain why different values are best for different systems.

An important concern in any simulation is how faithfully the true conformational dynamics of the complex are reproduced by the simulations. In some of the simulations, the ligands deviate significantly from the crystal structure. The most likely explanation is that the energy function (the MM force field and/or the solvation model) is inadequate. Even errors of a couple of kcal/mol can favor one binding mode over another. Protein-ligand complexes are a more strict test for force fields because the interactions are entirely nonbonded, and there are no constraints from chain connectivity and secondary structure. The deficiencies of explicit solvent simulations in modeling protein-ligand complexes may be less apparent because any deviations from the crystal structure take much longer to materialize (≈10 times longer than in implicit solvent simulations) because of the friction and caging effects of the solvent.[53] In several explicit solvent simulations of biotin-streptavidin, we found dome deviations from the crystal structure, although smaller than in the implicit solvent ones, possibly because of the shorter time scales covered. To prevent shifts from the crystal structure, one might suggest the use of soft harmonic constraints. However, this would prevent protein reorganization upon ligand binding. In some cases, crystal structures may be available for both the complex and the unbound protein, but small differences in resolution or refinement procedure of the crystals could produce artifacts when an energy difference is taken.

The extent of conformational fluctuations of ligands in protein binding sites is not known in detail. The fact that the ligands are visible in crystal structures suggests that they have a well-defined conformation. However, there is some evidence that ligand binding is not as static as it appears in crystal structures. Deng et al.[78] used temperature-jump relaxation spectroscopy with time-resolved fluorescence measurements to study the binding of NADH to lactate dehydrogenase and concluded that there are multiple, structurally different bound states at equilibrium, although these are not visible in the crystal structure. Similar observations were reported for NADP+ binding to dihydrofolate reductase using NMR.[79] Ladbury et al.[80] found that the state of a phosphotyrosine peptide bound to a SH2 domain is dynamic. The pY residue is fixed in its binding pocket, but the rest of the peptide is very mobile. Sometimes multiple binding modes are visible in crystal structures as well.[81–84] Brem and Dill[85] examined the consequences of multiple ligand binding modes for the calculation of binding affinities by using an exactly solvable lattice model. Our observations of multiple binding modes may actually reflect considerable flexibility of these ligands in the binding site of (strept)avidin.

Our results have implications for docking studies. Docking is usually performed with the protein kept rigidly in the conformation it has in the complex one is trying to reproduce. In a truly predictive setting, however, one would know the structure of the unbound protein or the structure of a complex with a different ligand. It is well known that docking is less successful in this case (e.g., see Ref. 86). Scoring of the putative structures of the complexes is done by more or less empirical functions,[87] which describe the interaction between ligand and protein without considering protein and ligand reorganization energy. It is conceivable that the protein reorganization energy is not the same for even slightly different ligands. For example, introduction of a polar group into a ligand may lead to a compensating loss of an intraprotein polar interaction and thus much less gain in affinity than one might expect. Although some studies of docking with a flexible receptor have been reported,[88] the emphasis so far has been on the effect of structural changes on the interaction with the ligand and not on the energetic consequences of the structural changes themselves. It would be interesting to examine whether consideration of protein reorganization energy improves the identification of binding sites in docking studies.

Future studies should focus on further characterization of protein and ligand reorganization in different systems and with different force fields and implicit solvation models. Further improvement of these energy functions will lead to more faithful simulations and more accurate binding thermodynamics, which, in conjunction with longer simulations and more systematic exploration of conformational space, should lead to improvements in binding affinity predictions.

ACKNOWLEDGMENTS

Useful comments by Dr. Stefan Boresch are gratefully acknowledged.

REFERENCES

1. Tembe B, McCammon JA. Ligand-receptor interactions. Comput Chem 1984;8:281–283.
2. Fleischman SH, Brooks CL III. Protein-drug interactions: characterization of inhibitor binding in complexes of DHFR with trimethoprim and related derivatives. Proteins 1990;7:52–61.
3. Lau FTK, Karplus M. Molecular recognition in proteins: simulation analysis of substrate binding by a tyrosyl-tRNA synthetase mutant. J Mol Biol 1994;236:1049–1066.
4. Bash PA, Singh UC, Brown FK, Langridge R, Kollman PA. Calculation of relative change in binding free energy of a protein-inhibitor complex. Science 1987;235:574–576.
5. Merz KM Jr, Murcko MA, Kollman PA. Inhibition of carbonic anhydrase. J Am Chem Soc 1991;113:4484–4490.
6. Essex JW, Severance DL, Tirado-Rives J, Jorgensen WL. Monte Carlo simulations for proteins: binding affinities for trypsin-benzamidine complexes via free-energy perturbations. J Phys Chem B 1997;101:9663–9669.
7. Singh SB, Ajay, Wemmer DE, Kollman PA. Relative binding affinities of distamycin and its analog to d(CGCAAGTTGGC)-*d(GCCAACTTGCG): comparison of simulation results with experiment. Proc Natl Acad Sci USA 1994;91:7673–7677.
8. Hermans J, Wang L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding: application to a complex of benzene and mutant T4 lysozyme. J Am Chem Soc 1997;119:2707–2714.
9. Helms V, Wade RC. Computational alchemy to calculate absolute protein-ligand binding free energy. J Am Chem Soc 1998;120:2710–2713.
10. Roux B, Nina M, Pomes R, Smith JC. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. Biophys J 1996;71:670–681.
11. Froloff N, Windemuth A, Honig B. On the calculation of binding

free energies using continuum methods: application to MHC class I protein-peptide interactions. Protein Sci 1997;6:1293–1301.
12. Polticelli F, Ascenzi P, Bolognesi M, Honig B. Structural determinants of trypsin affinity and specificity for cationic inhibitors. Protein Sci 1999;8:2621–2629.
13. Head MS, Given JA, Gilson MK. "Mining minima": direct computation of conformational free energy. J Phys Chem A 1997;101: 1609–1618.
14. Luo R, Gilson MK. Synthetic adenine receptors: direct calculation of binding affinity and entropy. J Am Chem Soc 2000;122:2934–2937.
15. Mardis KL, Luo R, Gilson MK. Interpreting trends in the binding of cyclic ureas to HIV-1 protease. J Mol Biol 2001;309:507–517.
16. Wang J, Szewczuk Z, Yue S-Y, Tsuda Y, Konishi Y, Purisima EO. Calculation of relative binding free energies and configurational entropies: a structural and thermodynamic analysis of the nature of nonpolar binding of thrombin inhibitors based on hirudin. J Mol Biol 1995;253:473–492.
17. Chong LT, Duan Y, Wang L, Massova I, Kollman PA. Molecular dynamics and free energy calculations applied to affinity maturation in antibody 48G7. Proc Natl Acad Sci USA 1999;96:14330–14335.
18. Kuhn B, Kollman PA. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. J Med Chem 2000;43:3786–3791.
19. Reyes CM, Kollman PA. Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. J Mol Biol 2000;297:1145–1158.
20. Noskov SY, Lim C. Free energy decomposition of protein-protein interactions. Biophys J 2001;81:737–750.
21. Baginski M, Fogolari F, Briggs JM. Electrostatic and non-electrostatic contributions to the binding free energies of antracycline antibiotics to DNA. J Mol Biol 1997;274:253–267.
22. Jayaram B, McConnell KJ, Dixit SB, Beveridge DL. Free energy analysis of protein-DNA binding: the EcoRI endonuclease-DNA complex. J Comput Phys 1999;151:333–357.
23. Bardi JS, Luque I, Freire E. Structure-based thermodynamic analysis of HIV-1 protease inhibitors. Biochemistry 1997;36:6588–6596.
24. Vajda S, Weng Z, Rosenfeld R, Delisi C. Effect of conformational flexibility and solvation on receptor-ligand binding free energies. Biochemistry 1994;33:13977–13988.
25. Åqvist J, Medina C, Samuelson J-E. A new method for predicting binding affinity in computer-aided drug design. Protein Eng 1994;7:385–391.
26. Hansson T, Åqvist J. Estimation of binding free energies for HIV proteinase inhibitors by MD simulations. Protein Eng 1995;8:1137–1144.
27. Wang J, Dixon R, Kollman PA. Ranking ligand binding affinities with avidin: a MD-based interaction energy study. Proteins 1999;34:69–81.
28. Wang W, Wang J, Kollman PA. What determines the van der Waals coefficient β in the LIE method to estimate binding free energies from MD simulations? Proteins 1999;34:395–402.
29. Smith RH Jr, Jorgensen WL, Tirado-Rives J, Lamb ML, Janssen PAJ, Michejda CJ, Smith MBK. Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using MC simulations in a linear response method. J Med Chem 1998;41:5272–5286.
30. Lamb ML, Tirado-Rives J, Jorgensen WL. Estimation of the binding affinities of FKBP12 inhibitors using a linear response method. Bioorg Med Chem 1999;7:851–860.
31. Sham YY, Chu ZT, Tao H, Warshel A. Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, PDLD/S-LRA calculations of ligands binding to an HIV protease. Proteins 2000;39:393–407.
32. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
33. Green NM. Avidin. Adv Protein Chem 1975;29:85–133.
34. Wilchek M, Bayer EA. Introduction to avidin-biotin technology. Methods Enzymol. 1990;184:5–13.
35. Miyamoto S, Kollman PA. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. Proteins 1993;16:226–245.
36. Miyamoto S, Kollman PA. What determines the strength of noncovalent association of ligands to proteins in aqueous solution? Proc Natl Acad Sci USA 1993;90:8402–8406.
37. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 1997;72:1047–1069.
38. Dixit SB, Chipot C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin-streptavidin system revisited. J Phys Chem A 2001;105:9795–9799.
39. Dixon RW, Kollman P. The free energies for mutating S27 and W79 to Ala in streptavidin and its biotin complex: the relative size of polar and nonpolar free energies on biotin binding. Proteins 1999;36:471–473.
40. Kuhn B, Kollman PA. A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning. J Am Chem Soc 2000;122:3909–3916.
41. Grubmüller H, Heymann B, Tavan P. Ligand-binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science 1996;271:997–999.
42. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys J 1997;72:1568–1581.
43. Finkelstein AV. Janin J. The price of lost freedom: entropy of bimolecular complex formation. Protein Eng 1989;3:1–3.
44. Chothia C, Janin J. Principles of protein–protein recognition. Nature 1975;256:705–708.
45. Tidor B, Karplus M. The contribution of vibrational entropy to molecular association. J Mol Biol 1994;238:405–414.
46. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy minimization and dynamics calculations. J Comput Chem 1983;4: 187–217.
47. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. J Chem Phys 1996;105:1902–1921.
48. Privalov PL, Makhatadze GI. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. J Mol Biol 1003;232:660–679.
49. Lazaridis T, Karplus M. Heat capacity and compactness of denatured proteins. Biophys Chem 1999;78:207–217.
50. Lazaridis T, Karplus M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. J Mol Biol 1999;288:477–487.
51. Dinner AR. Lazaridis T, Karplus M. Understanding β-hairpin folding. Proc Natl Acad Sci USA 1999;96:9068–9073.
52. Paci E, Smith LJ, Dobson CM, Karplus M. Exploration of partially unfolded states of human a-lactalbumin by MD simulations. J Mol Biol 2001;306:329–347.
53. Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
54. Inuzuka Y, Lazaridis T. On the unfolding of α-lytic protease and the role of the pro region. Proteins 2000;41:21–32.
55. Paci E, Karplus M. Unfolding proteins by external forces and temperature: the importance of topology and energetics. Proc Natl Acad Sci USA 2000;97:6521–6526.
56. Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased MD simulations. J Mol Biol 1999;288:441–449.
57. Kumar S, Sham YY, Tsai C-J, Nussinov R. Protein folding and function: the N-terminal fragment in adenylate kinase. Biophys J 2001;80:2439–2454.
58. Levy Y, Jortner J, Becker OM. Solvent effects on the energy landscape and folding kinetics of polyalanine. Proc Natl Acad Sci USA 2001;98:2188–2193.
59. Weber PC, Ohlendorf DH, Wendoloski JJ, Salemme FR. Structural origins of high-affinity biotin binding to streptavidin. Science 1989;243:85–88.
60. Pugliese L, Coda A, Malcovati M, Bolognesi M. Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2.7Å resolution. J Mol Biol 1993;231:698–710.
61. Katz BA. Binding to protein targets of peptidic leads discovered by phage display: crystal structures of streptavidin-bound linear and cyclic peptide ligands containing the HPQ sequence. Biochemistry 1995;34:15421–15429.

62. Goldstein H. Classical mechanics. 2nd ed. Reading: Addison-Wesley; 1980.
63. Pickett SD, Sternberg MJE. Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol 1993;231:825–839.
64. Nicholls A, Sharp KA, Honig B. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
65. Creamer TP. Side-chain conformational entropy in protein unfolded states. Proteins 2000;40:443–450.
66. D'Aquino JA, Freire E, Amzel LM. Binding of small organic molecules to macromolecular targets: evaluation of conformational entropy changes. Proteins Suppl 2000;4:93–107.
67. Brooks CL III, Karplus M. Deformable stochastic boundaries in molecular dynamics. J Chem Phys 1983;79:6312–6325.
68. Weber PC, Pantoliano MW, Thompson LD. Crystal structure and ligand binding studies of a screened peptide complexed with streptavidin. Biochemistry 1992;31:9350–9354.
69. Lazaridis T, Archontis G, Karplus M. Enthalpic contributions to protein stability: insights from atom-based calculations and statistical mechanics. Adv Protein Chem 1995;47:231–306.
70. Horton N, Lewis M. Calculation of the free energy of association for protein complexes. Protein Sci 1992;1:169–181.
71. Murphy KP, Xie D, Thompson KS, Amzel LM, Freire E. Entropy in biological binding processes: estimation of translational entropy loss. Proteins 1994;18:63–67.
72. Zidek L, Novotny MV, Stone MJ. Increased protein backbone conformational entropy upon hydrophobic ligand binding. Nat Struct Biol 1999;6:1118–1121.
73. Loh AP, Pawley N, Nicholson LK, Oswald RE. An increase in side chain entropy facilitates effector binding: NMR characterization of the side chain methyl group dynamics in Cdc42Hs. Biochemistry 2001;40:4590–4600.
74. Freitag S, Trong IL, Klumb L, Stayton PS, Stenkamp RE. Structural studies of the streptavidin binding loop. Protein Sci 1997;6:1157–1166.
75. Bostrom J, Norrby P-O, Liljefors T. Conformational energy penalties of protein-bound ligands. J Comp-Aided Mol Des 1998;12:383–396.
76. Vieth M, Hirst JD, Brooks CL III. Do active site conformations of small ligands correspond to low free energy solution structures? J Comp-Aided Mol Des 1998;12:563–572.
77. Mallik B, Masunov A, Lazaridis T. Distance and exposure dependent effective dielectric function. J Comp Chem. Submitted for publication.
78. Deng H, Zhadin N, Callender R. Dynamics of protein ligand binding on multiple time scales: NADH binding to lactate dehydrogenase. Biochemistry 2001;40:3767–3773.
79. Birdsall B, Feeney J, Tendler SJB, Hammond SJ, Roberts GCK. Dihydrofolate reductase: multiple conformations and alternative modes of substrate binding. Biochemistry 1989;28:2297–2305.
80. Ladbury JE. Hensmann M, Panayotou G, Campbell ID. Alternative modes of tyrosyl phosphopeptide binding to a Src family SH2 domain: implications for regulation of tyrosine kinase activity. Biochemistry 1996;35:11062–11069.
81. Zhou G, Ferrer M, Chopra R, Kapoor RM, Strassmaier T, Weissenhorn W, Skehel JJ, Oprian D, Schreiber SL, Harrison SC, Wiley DC. The structure of an HIV-1 specific cell entry inhibitor in complex with the HIV-1 gp41 trimeric core. Bioorg Med Chem 2000;8:2219–2228.
82. Montfort WR, Perry KM, Fauman EB, Finer-Moore JS, Maley GF, Hardy L, Maley F, Stroud RM. Structure, multiple site binding, and segmental accommodation in thymidylate synthase on binding dUMP and an anti-folate. Biochemistry 1990;29:6964–6977.
83. Murthy KHM, Winborne EL, Minnich MD, Culp JS, Debouck C. The crystal structure at 2.2 Å resolution of hydrosyethylene-based inhibitors bound to HIV-1 protease show that the inhibitors are present in two distinct orientations. J Biol Chem 1992;267:22770–22778.
84. Wojtczak A, Luft J, Cody V. Mechanism of molecular recognition: structural aspects of 3,3′-diiodo-L-:Thyronine binding to human serum transthyretin. J Biol Chem 1992;267:353–357.
85. Brem R, Dill KA. The effect of multiple binding modes on empirical modeling of ligand docking to proteins. Protein Sci 1999;8:1134–1143.
86. Murray CW, Baxter CA, Frenkel AD. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. J Comp-Aided Mol Des 1999;13:547–562.
87. Gohlke H, Klebe G. Statistical potentials and scoring functions applied to protein-ligand binding. Curr Opin Struct Biol 2001;11:231–235.
88. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. Mol. Pharmacol 2000;57:213–218.
89. Weber PC, Wendoloski JJ, Pantoliano MW, Salemme FR. Crystallographic and thermodynamic comparison of natural and synthetic ligands bound to streptavidin. J Am Chem Soc 1992;114:3197–3200.
90. Green NM. Avidin and streptavidin. Methods Enzymol 1990;184:51–67.