J|A|C|S
ARTICLES
Published on Web 05/28/2004

Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method

Donald Hamelberg* and J. Andrew McCammon‡

Contribution from the Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, and Department of Pharmacology, University of California, San Diego, La Jolla, California 92093-0365

Received August 7, 2003; E-mail: dhamelbe@mccammon.ucsd.edu

Abstract: Localized water molecules in the binding pockets of proteins play an important role in noncovalent association of proteins and small drug compounds. At times, the dominant contribution to the binding free energy comes from the release of localized water molecules in the binding pockets of biomolecules. Therefore, to quantify the energetic importance of these water molecules for drug design purposes, we have used the double-decoupling approach to calculate the standard free energy of tying up a water molecule in the binding pockets of two protein complexes. The double-decoupling approach is based on the underlying principle of statistical thermodynamics. We have calculated the standard free energies of tying up the water molecule in the binding pockets of these complexes to be favorable. These water molecules stabilize the protein−drug complexes by interacting with the ligands and binding pockets. Our results offer ideas that could be used in optimizing protein−drug interactions, by designing ligands that are capable of targeting localized water molecules in protein binding sites. The resulting free energy of ligand binding could benefit from the potential free energy gain accompanying the release of these water molecules. Furthermore, we have examined the theoretical background of the double-decoupling method and its connection to the molecular dynamics thermodynamic integration techniques.

Introduction

The solvent environment around biomolecules controls their structures and biological functions and plays important roles in protein−small molecule interactions. Association of small organic compounds to biomolecules is of fundamental importance to drug design; therefore, computational methods that can be used to study these interactions are of great importance. The release of water molecules and ions upon binding of ligands to protein binding sites contributes to changes in both entropy and enthalpy and at times is the dominant contribution to the binding free energy. A positive entropy change and, hence favorable free energy contribution, can arise from highly structured water molecules in the binding pocket that are released to a less ordered state in the bulk solvent upon binding. However, in some cases water molecules that are not released may take part in the interactions between drug molecules and biomolecules. These water molecules are often highly structured and localized; they may form water bridges between the drug molecules and the binding sites and optimize the interactions of the drug molecules. The release of localized water molecules may contribute favorably to the free energy of binding; therefore, they can be a target for drug design purposes.[1,2] A ligand that is designed to dislodge localized water molecules may have a more negative free energy of binding due to entropic gain than one that simply interacts and restricts them to the binding site. Experimentally,[1] it was shown that displacement of a localized water molecule in the binding pocket of the HIV-1/KNI-272 complex[3] by an inhibitor similar to KNI-272 causes the new inhibitor to bind more strongly than KNI-272. Therefore, the release of this water molecule appears to contribute favorably to the binding free energy of the inhibitor. This observation raises several questions concerning the free energy cost of tying up such a water molecule in the binding site of a protein and how this knowledge can be used in the study of noncovalent protein interactions.

Several theoretical studies[4−8] have been carried out on localized water molecules in and around protein structures, including one of the first applications of statistical thermodynamic perturbation techniques by Wade et al.[6] to such a system. Several groups[5,8] have subsequently used similar free energy

‡ Department of Pharmacology.

(1) Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C.-H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. *Science* **1994**, *263*, 380−384.
(2) Hodge, C. N.; Straatsma, T. P.; McCammon, J. A.; Wlodawer, A. In *Structural Biology of Viruses*; Chiu, W., Burnett, R. M., Garcea, R., Eds.; Oxford University Press: New York, 1997; pp 451−473.
(3) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. *Structure* **1995**, *3*, 581−590.
(4) Brooks, C. L., III; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159−181.
(5) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670−681.
(6) Wade, R. C.; Mazor, M. H.; McCammon, J. A.; Quiocho, F. A. *J. Am. Chem. Soc.* **1990**, *112*, 7057−7059.

10.1021/ja0377908 CCC: $27.50 © 2004 American Chemical Society

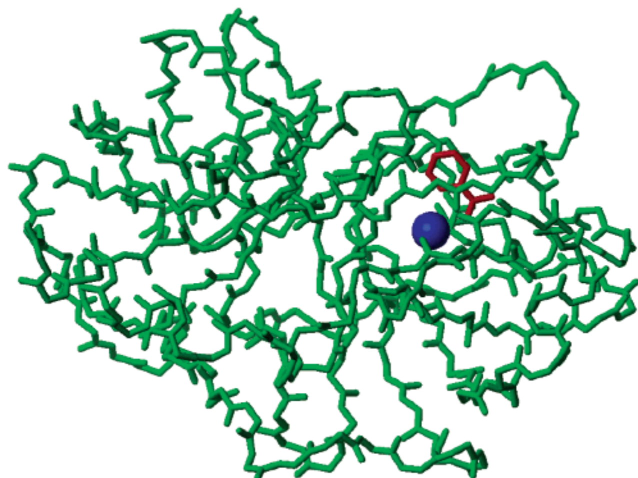

Figure 1. X-ray crystal structure of the trypsin (green)/benzyldiamine (red) complex with a bound water molecule (blue sphere) in the binding pocket.

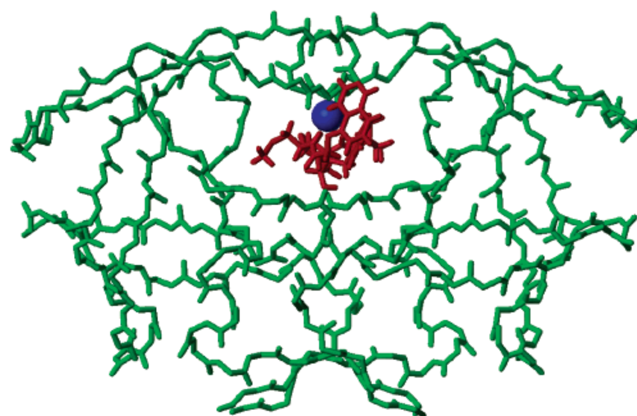

Figure 2. X-ray crystal structure of the HIV-1 (green)/KNI-272 (red) complex with a bound water molecule (blue sphere) in the binding pocket interacting with the inhibitor and the flaps of the protein.

methodology to study the properties of protein cavities. Zhang and Hermans[8] studied the hydrophilic nature of protein cavities by calculating the free energies of introducing a water molecule into these cavities using molecular dynamics free energy simulations. They were able to distinguish empty cavities from hydrated ones on the basis of the calculated free energies. Similarly, Roux et al.[5] used molecular dynamics free energy perturbation methods to calculate the stability of water molecules in the hydrophobic bacteriorhodopsin proton channel. Their results suggested that the transfer of four water molecules from bulk solvent to the channel is thermodynamically feasible and thereby shed some light on the mechanism of proton transfer in bacteriorhodopsin. Alternatively, recent work by Zheng and Lazaradis[9] used statistical mechanics equations of inhomogeneous fluid solvation theory to study the contribution of a water molecule to the energy, entropy, and heat capacity of protein solvation. Zheng and Lazaradis[9] studied the water molecule in the binding pocket of the HIV-1/KNI-272 protease inhibitor complex that interacts with the inhibitor and the flaps of the protein. They calculated the entropy loss in tying up this water molecule to be 9.8 cal mol^{-1} K^{-1} and the total contribution of this water molecule to the free energy of solvation to be −15.2 kcal/mol. Conversely, using entropic data from anhydrous salt and their corresponding hydrates, Dunitz estimated[10] the entropic cost of tying up a water molecule into a cavity of a crystalline protein to not exceed 7 cal mol^{-1} K^{-1}.

There has not been any attempt to quantitatively calculate the free energy contribution of bound water molecules to noncovalent protein−drug interaction using rigorous statistical mechanics molecular dynamics free energy simulation. In this work, we sought out to use the double-decoupling method[11] that was derived using the underlying theory of statistical mechanics as a framework to calculate the standard free energy of tying up water molecules in the binding site of protein complexes. Two systems in which bound water molecules are involved directly with the protein−ligand interaction are studied. The first system is the anionic trypsin wild-type protein.[12] The X-ray crystal structure (PDB ID 1ane) of trypsin complexed with benzyldiamine has an atomic resolution of 2.2 Å and shows an ordered water molecule donating two hydrogen bonds to TRP 190 and VAL 201 and accepting one hydrogen bond from the benzyldiamine ligand (Figure 1). The second system is a 2.0 Å atomic resolution X-ray crystal structure (PDB ID 1hpx) of HIV-1 protease complexed with KNI-272 inhibitor.[3] The crystal structure of the HIV-1 inhibitor complex shows an ordered water molecule donating two hydrogen bonds to KNI-272 and accepting two hydrogen bonds from ILE 50 and ILE 150 that are part of the protein flaps (Figure 2).

Theoretical Considerations

Background. The chemical potential of a molecule, A, in solution is given by

$$\mu_{\mathrm{sol,A}} = \mu^{\circ}_{\mathrm{sol,A}} + RT \ln\frac{\gamma_{\mathrm{A}} C_{\mathrm{A}}}{C^{\circ}} \tag{1}$$

where $\mu^{\circ}_{\mathrm{sol,A}}$ is the standard chemical potential, C_{A} is the concentration, and γ_{A} is the activity coefficient of A. R is the gas constant, T is the temperature, and C° is the standard concentration in the same units as C_{A}. Considering the noncovalent association of a ligand B and biomolecule A in aqueous solution as represented in eq 2,

$$\mathrm{A} + \mathrm{B} \Leftrightarrow \mathrm{AB} \tag{2}$$

the system is in equilibrium when

$$\mu_{\mathrm{sol,A}} + \mu_{\mathrm{sol,B}} = \mu_{\mathrm{sol,AB}} \tag{3}$$

Therefore, it can be shown that the standard free energy of binding is given by eq 4.

$$\Delta G^{\circ}_{\mathrm{AB}} = \mu^{\circ}_{\mathrm{sol,AB}} - (\mu^{\circ}_{\mathrm{sol,A}} + \mu^{\circ}_{\mathrm{sol,B}}) = -RT \ln K_{\mathrm{AB}} \tag{4}$$

where K_{AB} is the binding constant.

(7) Wade, R. C.; Mazor, M. H.; McCammon, J. A.; Quiocho, F. A. *Biopolymers* **1991**, *31*, 919−931.
(8) Zhang, L.; Hermans, J. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 433−438.
(9) Zheng, L.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 6636−6637.
(10) Dunitz, J. D. *Science* **1994**, *264*, 670.
(11) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(12) McGrath, M. E.; Vasquez, J. R.; Craik, C. S.; Yang, A. S.; Honig, B.; Fletterick, R. J. *Biochemistry* **1992**, *31*, 3059−3064.

The statistical mechanics expression[13] for of the standard chemical potential $(\partial G/\partial n_A)_{T,P}$ of A in solution is given by eq 5:

$$\mu^\circ_{\mathrm{sol,A}} = -RT\ln\left(\frac{1}{V_{\mathrm{N,A}}C^\circ}\frac{Q_{\mathrm{N,A}}(V_{\mathrm{N,A}})}{Q_{\mathrm{N,0}}(V_{\mathrm{N,0}})}\right) + P^\circ\overline{V}_{\mathrm{A}} \quad (5)$$

Here $Q_{N,A}(V_{N,A})$ is the canonical partition function for a system containing a large number of N solvent molecules and one solute molecule A at volume $V_{N,A}$. $V_{N,A}$ is the volume of A when it is at equilibrium at standard pressure P°. $Q_{N,0}(V_{N,0})$ is the partition function of the solvent, without the solute molecule, at equilibrium volume $V_{N,0}$. Assuming that classical statistical mechanics apply to this system, the ratio of the partition functions[14] in eq 5 is depicted in eq 6 in terms of the Hamiltonian, H.

$$\frac{Q_{\mathrm{N,A}}}{Q_{\mathrm{N,0}}} = \frac{\int \mathrm{d}\mathbf{p}_{\mathrm{A}}\,\mathrm{d}\mathbf{p}_{\mathrm{S}}\int \mathrm{d}\mathbf{r'}_{\mathrm{A}}\,\mathrm{d}\mathbf{r}_{\mathrm{S}}\exp[-\beta H(\mathbf{p}_{\mathrm{A}},\mathbf{p}_{\mathrm{S}},\mathbf{r'}_{\mathrm{A}},\mathbf{r}_{\mathrm{S}})]}{\sigma_{\mathrm{A}}\int \mathrm{d}\mathbf{p}_{\mathrm{S}}\int \mathrm{d}\mathbf{r}_{\mathrm{S}}\exp[-\beta H(\mathbf{p}_{\mathrm{S}},\mathbf{r}_{\mathrm{S}})]} \quad (6)$$

where $\beta = (RT)^{-1}$, σ_A is the symmetry number of the solute A, and H of the system is in terms of the conjugate momenta and coordinates as given by eq 7:

$$H(\mathbf{p}_{\mathrm{A}},\mathbf{p}_{\mathrm{S}},\mathbf{r'}_{\mathrm{A}},\mathbf{r}_{\mathrm{S}}) = \sum_{i=1}^{M_{\mathrm{A}}+M_{\mathrm{S}}}\frac{p_i^2}{2m_i} + U(\mathbf{r'}_{\mathrm{A}},\mathbf{r}_{\mathrm{S}}) \quad (7)$$

Here M_A and M_S, respectively, are the numbers of atoms of the one solute molecule and the solvent molecules, $\mathbf{r'}_A$ represents the internal and external coordinates of A, and $\mathbf{r}_S$ represents the coordinates of the solvent. The coordinate system of A can be separated into internal, $\mathbf{r}_A$, and external, $\varsigma_A = (\mathbf{R}_A,\psi,\theta,\varphi)$, coordinates.[15] The external coordinates can be defined as the three degrees of freedom due to rotation as defined by the three Euler angles (ψ, θ, and φ) of any three atoms used to define the molecular axes of *A*, and three translational degrees of freedom corresponding to the position of the first atom or center of mass of A, $\mathbf{R}_A$. Equation 5 can therefore be written as:

$$\mu^\circ_{\mathrm{sol,A}} = -RT\ln\left(\frac{8\pi^2}{C^\circ\sigma_{\mathrm{A}}}\prod_{i=1}^{M_{\mathrm{A}}}(2\pi m_iRT)^{3/2}\frac{Z_{\mathrm{N,A}}}{Z_{\mathrm{N,0}}}\right) + P^\circ\bar{V}_{\mathrm{A}} \quad (8)$$

where $Z_{N,A}$ and $Z_{N,0}$ are the configuration integrals given by eq 9 and 10. The factor $8\pi^2$ is due to the rotational part of the integral over the external coordinates ζ_A, which is equal to $8\pi^2V_{N,A}$. The volume $V_{N,A}$ cancels out. $(2\pi m_iRT)^{3/2}$ is the classical approximation of the integral over momentum.

$$Z_{\mathrm{N,A}} = \int \exp[-\beta U(\mathbf{r}_{\mathrm{A}},\mathbf{r}_{\mathrm{S}})]\,\mathrm{d}\mathbf{r}_{\mathrm{A}}\,\mathrm{d}\mathbf{r}_{\mathrm{S}} \quad (9)$$

$$Z_{\mathrm{N,0}} = \int \exp[-\beta U(\mathbf{r}_{\mathrm{S}})]\,\mathrm{d}\mathbf{r}_{\mathrm{S}} \quad (10)$$

Therefore, from eqs 4 and 8, the standard free energy of binding can be written in terms of the configuration integrals:

$$\Delta G^\circ_{\mathrm{AB}} = -RT\ln\left(\frac{C^\circ}{8\pi^2}\frac{\sigma_{\mathrm{A}}\sigma_{\mathrm{B}}}{\sigma_{\mathrm{AB}}}\frac{Z_{\mathrm{N,AB}}Z_{\mathrm{N,0}}}{Z_{\mathrm{N,A}}Z_{\mathrm{N,B}}}\right) + P^\circ\overline{V}_{\mathrm{AB}} \quad (11)$$

Double-Decoupling Method. The thermodynamics breakdown shown in Figure 3 and the statistical mechanics analysis of the double-decoupling method presented below are taken from Gilson et al.[11] with similar notations. The article by Gilson et al. serves as a thorough description of the double decoupling methodology. Calculation of the standard free energy, ΔG_2°, of removing B from the bulk solvent to the gas phase as shown in Figure 3 is straightforward, since it does not depend on the choice of the standard concentration; ΔG_2° can be determined by simply removing B from solution using molecular dynamics free energy perturbation simulation.

However, calculating the standard free energy of transferring B from the complex to the gas phase, ΔG_1°, is not that straightforward. ΔG_1° is defined as the free energy of decoupling B from the binding site of A during the simulation. Nevertheless, during the later part of the simulation when B is weakly coupled to A, B would have to explore the entire simulation box in order for ΔG_1° to converge. This can be circumvented by constraining the coordinates of B to occupy the binding site of A using a restraining potential while decoupling the interactions of B. At the end of decoupling B from A, B is simply an ideal gas constrained in the binding site with a definite chemical potential. Therefore, ΔG_1° can be express as:

$$\Delta G_1^\circ = (\mu^\circ_{\mathrm{sol,A}} + \mu^\circ_{\mathrm{gas,B}}) - \mu^\circ_{\mathrm{sol,AB}} \quad (12)$$

where $\mu^\circ_{\mathrm{gas,B}}$ is the standard chemical potential of B in the ideal gas phase and is given by eq 13:

$$\mu^\circ_{\mathrm{gas,B}} = -RT\ln\left(\frac{8\pi^2}{C^\circ\sigma_{\mathrm{B}}}\prod_{i=1}^{M_{\mathrm{A}}}(2\pi m_iRT)^{3/2}Z_{0,\mathrm{B}}\right) \quad (13)$$

Therefore, eq 12 becomes:

$$\begin{aligned}\Delta G_1^\circ = {} & -RT\ln\left(\frac{Z_{\mathrm{N,A}}Z_{0,\mathrm{B}}}{Z_{\mathrm{N,AB}}}\right) \\ & -RT\ln\left(\frac{\sigma_{\mathrm{AB}}}{\sigma_{\mathrm{A}}\sigma_{\mathrm{B}}}\right) \\ & +RT\ln(C^\circ) - RT\ln(8\pi^2) \\ & +P^o(\bar{V}_{\mathrm{A}} - \bar{V}_{\mathrm{AB}})\end{aligned} \quad (14)$$

Using previous equations and defining an artificial potential function $U(\lambda, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)$ that interpolates between the initial energy function ($\lambda = 0$, AB_{sol}) and the final energy function ($\lambda = 1$, $A_{sol} + B_{gas}$), we can write ΔG_1° as follows.

$$\begin{aligned}\Delta G_1^\circ = {} & \int\left\langle\frac{\partial U(\lambda,\mathbf{r}_{\mathrm{A}},\mathbf{r}_{\mathrm{B}},\zeta_{\mathrm{B}},\mathbf{r}_{\mathrm{S}})}{\partial\lambda}\right\rangle_\lambda \mathrm{d}\lambda \\ & -RT\ln\left(\frac{\sigma_{\mathrm{AB}}}{\sigma_{\mathrm{A}}\sigma_{\mathrm{B}}}\right) \\ & +RT\ln(C^oV_{\mathrm{I}}) + RT\ln(\xi_{\mathrm{I}}/8\pi^2) \\ & +P^o(\bar{V}_{\mathrm{A}} - \bar{V}_{\mathrm{AB}})\end{aligned} \quad (15)$$

(13) Hill, T. L. *Cooperativity Theory in Biochemistry*; Springer-Verlag: New York, 1985.
(14) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1973.
(15) Steinberg, I. Z.; Scheraga, H. A. *J. Biol. Chem.* **1963**, *238*, 172−181.

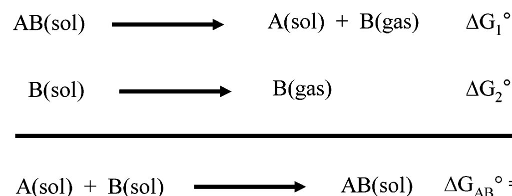

Figure 3. Thermodynamic breakdown of the double-decoupling method. The notations (sol) and (gas) imply a molecule in solution and in ideal gas phase, respectively.

where V_I is the integral over the volume element in which the ideal gas is constrained to occupy, and ξ_I is the integral over the rotational space the ideal gas is allowed to sample. Equation 15 is the same as eq 28 in Gilson et al.[11] The first term can be evaluated using thermodynamic integration[16] by molecular dynamics simulation as λ is varied from 0 to 1. 0 in the AB(sol) state and 1 is the A(sol) + B(gas) state. The third and fourth terms can be viewed as correction terms due to the constraining of B during the free energy simulation. They are the change in free energy when the constrained ligand is allowed to expand and occupy a volume of 1/C° and to rotate freely, respectively. Therefore, if the rotational space is represented by the three Euler angles, and B is allowed to rotate freely,[15] then ξ_I is given by eq 16 and is evaluated to be $8\pi^2$. This condition will reduce the fourth term of eq 14 to zero.

$$\xi_I = \int_{\theta=0}^{\pi} \sin\theta \, d\theta \int_{\psi=0}^{2\pi} d\psi \int_{\varphi=0}^{2\pi} d\varphi \quad (16)$$

Evaluation of V_I. We have chosen to use a harmonic potential to constrain the coordinates of B to occupy the binding site of A. To constrain the ligand B to occupy the binding site of A using a harmonic potential, it is necessary to carefully estimate the harmonic force constant, k. There is no absolute approach or an exact statistical mechanics prescription as to how the constraint should be applied. However, it is necessary for the harmonic constraint potential to be tuned to correspond to the range of motion of the ligand in the fully coupled simulation. This is necessary because if the constraint is too strong the region sampled by B will be restricted and will lead to inadequate sampling. By the same token, if the constraint potential is too weak, the simulation will fail to converge because the region the ligand B is constrained to occupy becomes so large.

The configuration integral of B(gas) first introduced in eq 13 is given by eq 17:

$$Z_{0,B} = \int \exp[-\beta U(\mathbf{r}_B)] \, d\mathbf{r}_B \quad (17)$$

For an ideal gas with no interaction where $U = 0$, the configuration integral $Z_{0,B}$ is thus equivalent to the volume B occupies. However, if B is constrained to occupy a region with a harmonic potential with force constant k, then the potential energy function becomes:

$$U(\mathbf{r}_B) = -\frac{k}{2}(r_B - r_0)^2 \quad (18)$$

Hence the volume element, V_I, in eq 15 which B is constrained to occupy is calculated to be:

$$V_I = \int_v \exp[-k/2RT(r_B - r_0)^2] \, d\mathbf{r}_B = \left(\frac{2\pi RT}{k}\right)^{3/2} \quad (19)$$

From eqs 17 and 18, the probability distribution in one dimension is given by:

$$P(x) = \frac{\exp[-k/2RT(x - x_0)^2]}{\int \exp[-k/2RT(x - x_0)^2] \, dx} \quad (20)$$

Therefore, the square positional fluctuation in the x-direction is given by eq 21:

$$\langle \delta x^2 \rangle = \langle (x - x_0)^2 \rangle = \int (x - x_0)^2 P(x) \, dx = \frac{RT}{k} \quad (21)$$

This is, of course, consistent with the equipartition theorem. Hence eq 22:

$$\langle \delta r^2 \rangle = \frac{3RT}{k} \text{ and } k = \frac{3RT}{\langle \delta r^2 \rangle} \quad (22)$$

Therefore, the force constant can be accurately estimated using eq 22 by calculating the atomic fluctuation, δr, of the constrained atom or center of mass of B during the course of a molecular dynamics simulation with the fully unperturbed potential function.

Computational Methods

Equilibration and Molecular Dynamics Simulation. The coordinates of the two model systems used in this study were from the crystal structures obtained from the protein data bank. The hydrogens were added to the heavy atoms using the Leap module in AMBER. Furthermore, the atomic charges and force field parameters of benzyldiamine and KNI-272 were determined using the RESP[17,18] methodology at the HF/6-31G* level of theory in Gaussian 98,[19] and comparable standard parameters of the Cornell et al. force field.[20] Each complex was solvated with TIP3P[21] water molecules to fill a periodic box. Eight Na^+ ions and 4 Cl^- ions were placed around the trypsin/benzyldiamine and HIV-1/KNI-272 complexes, respectively, using the Leap module in AMBER to obtain electrostatic neutrality. All molecular dynamics simulations were carried out with the AMBER 6 suites of program.[22]

(16) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.

(17) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(18) Cornell, W. D.; Cieplak, P.; Balyl, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.

(20) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(21) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(22) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *Amber 6*, 5.0 ed.; University of California: San Francisco, 1997.

The systems were equilibrated by using a multistage equilibration protocol. At the start of the equilibration, 500 kcal mol^{-1} Å^2 harmonic constraints were placed on the complex molecules. The water and ions were minimized for 1000 steps, followed by molecular dynamics for 25 ps. The equilibration was continued with a 25 ps MD run at 300 K with 100 kcal mol^{-1} Å^2 constraints placed on the protein complex. This was followed by a 1000 step minimization on the entire system with a 25 kcal mol^{-1} Å^2 constraints placed on the complex. Finally, the entire system was heated from 100 to 300 K over 10 ps. After equilibration, the final equilibrated structures were used to carry out 0.5 ns of MD simulations in the NPT ensemble with periodic boundary conditions at a constant temperature of 300 K with the Berendsen temperature algorithm[23] and at a pressure of 1 bar. The SHAKE[24] algorithm was applied to all bonds involving hydrogen atoms, and an integration time step of 2.0 fs was used in solving Newton's equation of motion. The particle mesh Ewald[25] method was used to treat the long-range electrostatic interactions, and the nonbonded van der Waals interactions were subjected to a 12 Å cutoff. The snapshots of the MD trajectories were collected every 1 ps. The resulting trajectories were used to calculate the atomic positional fluctuation of the oxygen atom of the bound water molecule in order to estimate the harmonic force constant according to eq 22. The calculated harmonic force constant was used to constrain the oxygen atom of the water molecule to occupy the binding site of the protein complex during the free energy simulation.

Double-Decoupling Free Energy Simulation. Our double-decoupling free energy simulations involve gradually turning off the electrostatic and van der Waals interactions of the water molecule from the rest of the system. This involves two sets of simulations: the transfer of the water molecule from bulk water to the gas phase, ΔG_2°, and the transfer of the water molecule from the binding pocket of the protein complex to the gas phase (part of ΔG_1°), during which the water molecule is constrained to occupy the binding site as defined by the coordinate system of the complex. In the second simulation, the water molecule is constrained to occupy the binding site using a harmonic potential with an appropriate force constant calculated using eq 22 as described above. To perform the change from H_2O_{sol} to H_2O_{gas}, the energy term is progressively mapped from U(sol) to U(gas) along a chosen path. This chosen path is mapped as a function of a coupling parameter λ that varies from 0 to 1, where $U = U(\lambda)$, $U(0) = U(\text{gas})$, and $U(1) = U(\text{sol})$. The free energy difference between the two states is calculated by the thermodynamic integration approach at discrete points of λ_i along the path using a fixed width for $\Delta\lambda$. A value of 0.01 for $\Delta\lambda$ was used for the transfer of the water molecule from bulk water to the gas phase, and 0.05 was used for transferring the water molecule from the binding pocket of the protein complex to the gas phase. The free energy change for each simulation was calculated by varying λ from 0 to 1(forward), and then from 1 to 0 (backward). The simulation for each value of λ (window) was initially equilibrated for 10.0 ps, and the data sampling was also performed for 10.0 ps, and both were progressively increased by 1.0 ps during subsequent runs until the hystereses were very small. The simulation conditions for the free energy simulation were similar to that of the molecular dynamics simulation. The final structure of the molecular dynamics simulation was used as the starting structure of the free energy simulation.

Results and Discussion

The standard free energy of tying up a water molecule in the binding site of a protein was studied using the double-decoupling method. This approach used in calculating the standard free energy of binding was derived from the underlying classical statistical mechanics theory and made the connection to molecular dynamics free energy perturbation techniques. A review by Gilson et al.[11] showed that attempts by several groups[26−30] to calculate the absolute free energy of binding fell short because the theory behind the methodologies used was not well grounded in statistical mechanics. The first part of this method, as shown in Figure 3, is the calculation of the free energy of decoupling the water molecule from bulk solvent into an ideal gas phase, ΔG_2°. The calculation of the standard free energy of removing a water molecule from bulk solvent into gas phase does not depend on the standard concentration and is therefore straightforward. This was performed using the molecular dynamics thermodynamic integration technique. ΔG_2° was calculated to be 6.0 kcal/mol as laid out in Table 1. Our calculated standard free energy of decoupling a water molecule from the bulk solvent is similar to an experimentally observed free energy of 6.3 kcal/mol[31] and previously calculated values of 6.1[32] and 6.4[33] kcal/mol. Table 1 also shows the separate contributions of the electrostatic and van der Waals interactions to the total standard free energy.

Table 1. Standard Free Energy Change in kcal/mol for Removing a Water Molecule from Bulk Solvent

	ΔG°_2		
	elec	vdW	total
forward	8.1	−2.2	5.9
reverse	8.2	−2.1	6.1
average	8.2 ± 0.1	−2.2 ± 0.1	6.0 ± 0.2

The second component of the double-decoupling method is the calculation of the standard free energy of decoupling the water molecule from the binding site of the protein complex. While decoupling the interactions of the water molecule from the protein environment, the coordinates of the oxygen atom on the water molecule were harmonically constrained to occupy the binding site of the complex as defined by the crystal structure. The atomic positional fluctuation of the oxygen atom and the harmonic force constant were calculated from the preceding molecular dynamics simulations for the trypsin and HIV-1 complexes. The positional fluctuation of the water molecule in the binding pocket of the trypsin complex is larger than that of the HIV-1 complex. This shows that the water molecule in the trypsin complex is slightly less localized than that of the HIV-1 complex. Therefore, the appropriate harmonic force constants necessary to constrain the water molecule in the binding site of the trypsin and HIV-1 complexes were calculated to be 6.6 and 16.4 kcal mol^{-1} A^{-2}, respectively. Tables 2 and 3 present all the terms (according to eq 15) that

(23) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A. *J. Chem. Phys* **1984**, *81*, 3684−3690.
(24) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(25) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(26) Miyamoto, S.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402−8406.
(27) Miyamoto, S.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 226−245.
(28) Pranata, J.; Jorgensen, W. L. *Tetrahedron* **1991**, *47*, 2491−2501.
(29) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742−3746.
(30) Sneddon, S. F.; Tobias, D. J.; Brooks, C. L., III. *J. Mol. Biol.* **1989**, *209*, 817−820.
(31) Bennaim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016−2027.
(32) Jorgensen, W. L.; Blake, J. F.; Buckner, J. K. *Chem. Phys.* **1989**, *129*, 193−200.
(33) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.

Table 2. The Standard Free Energy (kcal/mol) and Its Components of Tying Up a Water Molecule in the Binding Pocket of Trypsin/Benzylamine Complex

	$\Delta G°_1$						$\Delta G°_{AB}$
	$\int\langle\partial U/\partial\lambda\rangle_\lambda\, d\lambda$		$-RT\ln(\sigma_{AB}/\sigma_A\sigma_B)$	$RT\ln(C°V_I)^a$	$RT\ln(\xi_I/8\pi^2)$	total	$\Delta G°_2 - \Delta G°_1$
	elec	vdW					
forward	9.4	2.8	0.4	−4.9	0.0	7.7	
reverse	9.6	3.0	0.4	−4.9	0.0	8.1	
						7.9 ± 0.3	−1.9 ± 0.5

[a] The atomic fluctuation, $\delta r = 0.52$ Å, and the corresponding harmonic force constant, $k = 6.6$ kcal mol^{-1} Å^{-2}. F = forward and R = reverse.

Table 3. The Standard Free Energy (kcal/mol) and Its Components of Tying Up a Water Molecule in the Binding Pocket of HIV-1/KNI-272 Complex

	$\Delta G°_1$						$\Delta G°_{AB}$
	$\int\langle\partial U/\partial\lambda\rangle_\lambda\, d\lambda$		$-RT\ln(\sigma_{AB}/\sigma_A\sigma_B)$	$RT\ln(C°V_I)^a$	$RT\ln(\xi_I/8\pi^2)$	total	$\Delta G°_2 - \Delta G°_1$
	elec	vdW					
forward	12.9	1.7	0.4	−5.7	0.0	9.3	
reverse	12.5	1.6	0.4	−5.7	0.0	8.8	
						9.1 ± 0.4	−3.1 ± 0.6

[a] The atomic fluctuation, $\delta r = 0.33$ Å, and the corresponding harmonic force constant, $k = 16.4$ kcal mol^{-1} Å^{-2}. F = forward and R = reverse.

are required to calculate the standard free energy of tying up a water molecule into the binding site of the trypsin and HIV-1 complexes, respectively. The first term is the electrostatic and van der Waals contribution from the thermodynamic integration simulation, the second term takes care of the fact that the water molecule is symmetrical and has a symmetry number (σ_B) of 2, the third term is a correction because the translational degrees of freedom of the bound water molecule were constrained during the thermodynamic integration simulation, and the fourth term is zero because the bound water molecule was allowed to rotate freely during the free energy simulation. The standard free energies for tying up a water molecule in the binding pockets of both complexes are less than zero. This implies that the localized water molecules help in stabilizing the protein−ligand interactions of both complexes. Also, the results confirm the assignments of these water molecules in the binding site of the X-ray crystal structures. The free energy of tying up the water molecule for the HIV-1 complex is slightly lower than that of the trypsin complex. Therefore, the water molecule in the HIV-1 protease binding site is more tightly bound than that in the trypsin complex.

As prescribed by the double-decoupling approach, the ligand is constrained to occupy the binding site while the interactions of the ligand are being gradually decoupled from the binding site. However, there is no precise statistical mechanics definition of the retraining potential. We chose to use a harmonic restraining potential on the oxygen atom of the water molecule. To use a harmonic potential, it is important to accurately determine the required force constant, k. Therefore, this raises important questions as to how sensitive are the calculated free energies to the strength of the restraining potentials and whether an arbitrary choice of k would yield the correct results. In answering these questions, we carried out two additional free energy calculations on the HIV-1/KNI complex using a force constant 10 times stronger (164.0 kcal mol^{-1} Å^{-2}) and 10 times weaker (1.64 kcal mol^{-1} Å^{-2}) than the calculated force constant of 16.4 kcal mol^{-1} Å^{2}. The free energy of tying up the water molecule in the binding site of the HIV-1 complex was calculated to be −0.8 ± 0.1 kcal/mol using a restrictive restraining potential (force constant of 164.0 kcal mol^{-1} Å^{-2}). This value is significantly greater than the free energy when calculated with the force constant determined from the range of motion of the water molecule during the MD simulation. This difference can be attributed to the fact that at a much higher value of k the region occupied by the water molecule is further restricted, thus leading to inadequate sampling. On the other hand, the free energy at a very low value k was calculated to be −3.4 ± 1.5 kcal/mol. This value is comparable to the free energy determined with $k = 16.4$ kcal mol^{-1} Å^{2}. When k is very small, the restraining potential becomes weak, and the simulation fails to properly converge, as evident from the large error, because the region the water molecule is constrained to occupy becomes too large. Theoretically, with a weaker restraining potential, the correct free energy could be determined if the water molecule is allowed to sufficiently sample the region which it is constrained to occupy.

Our results clearly show the importance of localized water molecules in the binding pockets of proteins, and because of their energetic potential this makes them a target for drug design purposes. Releasing these highly structured water molecules will favorably contribute to the free energy of a ligand that can displace them. According to the rationale by Dunitz,[10] the entropic gain of releasing a bound water molecule for a protein can be as high as 7 cal mol^{-1} K^{-1}, which corresponds to a free energy gain of about 2.1 kcal/mol at 300 K. The enthalpic contribution to the total free energy will depend on the types of interactions experienced by the water molecules in the binding site. Therefore, it can be seen that the entropic contribution to the binding free energy by releasing water molecules and counterions from the binding sites of protein molecules is a very important component that cannot be ignored in drug design and in understanding noncovalent protein interactions. The results provided here give some indication of how ligands could be designed to increase their binding affinity. If a substituent is added to the ligand that displaces a bound water molecule and if the net contribution of this substituent is greater than the

modest free energy cost of displacing the solvent, the result will be an increase in ligand affinity. In practice, this may not be too difficult to do, because the ligand will already have paid the price of translational and rotational entropy loss. Its substituent need not pay this price again, but can harvest the entropy of releasing the water molecule.

Acknowledgment. This work was supported in part by grants from NSF, NIH, the Center for Theoretical Biological Physics, the National Biomedical Computation Resource, San Diego Supercomputing Center, and Accelrys, Inc.

JA0377908