JCTC Journal of Chemical Theory and Computation

Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force

Slimane Doudou, Neil A. Burton, and Richard H. Henchman*

School of Chemistry, The University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom, and Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom

Received June 20, 2008

Abstract: A practical approach that enables one to calculate the standard free energy of binding from a one-dimensional potential of mean force (PMF) is proposed. Umbrella sampling and the weighted histogram analysis method are used to generate a PMF along the reaction coordinate of binding. At each point, a restraint is applied orthogonal to the reaction coordinate to make possible the determination of the volume sampled by the ligand. The free energy of binding from an arbitrary unbound volume to the restrained bound form is calculated from the ratio of the PMF integrated over the bound region to that of the unbound. Adding the free energy changes from the standard-state volume to the unbound volume and from the restrained to the unrestrained bound state gives the standard free energy of binding. Exploration of the best choice of binding paths is also made. This approach is first demonstrated on a model binding system and then tested on the benzamidine−trypsin system for which reasonable agreement with experiment is found. A comparison is made with other methods to obtain the standard free energy of binding from the PMF.

Introduction

The binding free energy is a fundamental property of many molecular systems[1−9] such as those encountered in drug design, catalysis, and self-assembly. Consequently, the development of free energy methods plays a pivotal role in research. Typically, these methods can be used either to calculate the free energy of the bound and unbound states separately, in approaches such as the MM/PBSA[10−12] and LIE[1,13]methods, or to evaluate the free energy difference between bound and unbound states. Free energy difference methods proceed through decoupling the interactions between the ligand and its receptor, giving a nonphysical pathway[8,10,14,15] or by displacing the ligand along a physical pathway of binding.[10,12,16−21]Although the decoupling free energy methods are commonly used to calculate the binding free energy, the pathway methods also provide mechanistic and kinetic information on the real process of binding. The immediate output of a binding-pathway free energy method is not a free energy difference but a potential of mean force (PMF), which is defined as the negative logarithm of the probability of being at a given value of a specified reaction coordinate. Sampling realistic binding pathways is an ongoing challenge, and the path usually has to be selected and biased to achieve converged sampling, an exception to this being transition path sampling.[22]Many different pathway methods have been developed. These include umbrella sampling,[23] adaptive force bias,[24] the Jarzynski method,[25] and metadynamics.[26] While newer methods such as the adaptive force bias method should be more efficient, umbrella sampling has been one of the most widely used methods to study binding in solution.[12,16−20,27−30]

For protein−ligand systems, the relationship between the standard free energy of binding ΔG° and the change in the PMF on binding, hereafter referred to as the PMF depth, ΔW_R, is controversial, and they are often assumed to be equivalent.[31−34] Even in the few studies that examine that relationship, there are differing formulations.[10,12,18,21,30,35−37] In the simple case of spherical symmetry, the association force constant can be calculated by integrating the PMF over the bound region and by accounting for the radial Jacobian

* Corresponding author phone: +44 (0)161 306 5194; fax: +44 (0)161 306 5201; e-mail: henchman@manchester.ac.uk.

10.1021/ct8002354 CCC: $40.75
Published on Web 03/10/2009

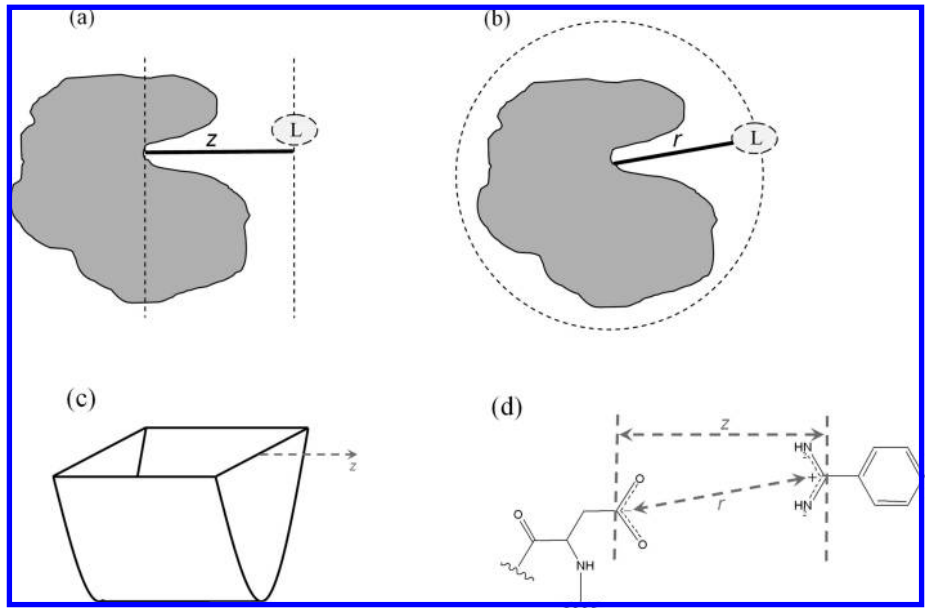

Figure 1. (a) *z*-Component reaction coordinate and (b) the distance *r* reaction coordinate between the protein on the left and the ligand on the right. (c) The three-dimensional PMF of the model binding site. (d) The *z*-component and the distance *r* reaction coordinates between the carboxylate of Asp 189 and the amidine carbon of benzamidine.

and standard-state concentration.[10,30,35−37] In protein−ligand systems, there are two major problems. First, there is no such symmetry due to the complex shapes of the proteins. Second, there are limitations on sampling since the ligand would need to sample all the space around the protein. This makes necessary more general formulations. Gilson and co-workers[10] suggest calculating the standard free energy of binding from the PMF by defining w_b as the average PMF depth in the bound region and adding to it a term that accounts for the change in the ligand's translational and rotational volumes upon binding. Their PMF is a six-dimensional function of the ligand's position and orientation. The problem with using a PMF of such high dimensionality is that the extensive sampling required makes it difficult to calculate. A similar approach was employed by Lee and Olson,[12,18] but they use w_{min}, the minimum of the PMF, instead of w_b and their PMF is a function of a one-dimensional reaction coordinate. Woo and Roux[21] have a different approach in which they apply multiple restraints to the ligand, determine a radial one-dimensional PMF, and then remove the restraints. The standard free energy of binding is then obtained from the equilibrium binding constant using a complex derivation that involves calculating a surface area integral for the ligand at a large distance from the receptor and multiplying this by terms to add and remove the restraints, by the PMF integral over the bound state, and by the standard concentration. Information about the PMF depth is contained implicitly in the PMF integral.

In this work, a strategy is described to obtain the standard free energy of binding from a one-dimensional PMF. Our approach is similar to Woo and Roux's[21] but is constructed in a simpler manner that permits an easier implementation and physical interpretation of the terms involved. The choice of reaction coordinate is important in formulating the method and achieving converged sampling. The PMF is obtained along a one-dimensional *z*-component reaction coordinate between the protein and the ligand while applying restraints orthogonal to this reaction coordinate. The protein is oriented so that the *z*-axis points directly out of the binding site. The orthogonal restraints serve a dual purpose: they make it possible to calculate the area sampled by ligand orthogonal to the reaction coordinate and they limit the sampling required in each window. The use of a *z*-component rather than a radial reaction coordinate *r* is illustrated in Figure 1a and 1b and ensures that the configurational areas at each point along the reaction coordinate are similar in size rather than increasing as $4\pi r^2$. Small, constant areas are easier to sample than areas that increase in size. An extra term accounting for the $4\pi r^2$ increase in area as the ligand unbinds would be required if a flat PMF is desired in the bulk solvent.[36,38,39] Computing the ratio of the integrated PMF over the bound and unbound regions leads to the free energy of binding from a specified unbound volume to a bound restrained state. The orthogonal restraint permits the calculation of the unbound volume. Adding the free energy changes from the standard volume to this unbound volume and from the restrained to the unrestrained bound state determines the standard free energy of binding. The approach is used on a model system to demonstrate its performance. It is then applied to calculate the standard Gibbs free energy of binding of trypsin to benzamidine along a suitably chosen binding path and then compared with experiment.

Methods

Standard Free Energy of Binding from the PMF. The reaction coordinate of the PMF is chosen to be the *z*-component.[40,41] Orthogonal harmonic restraints are applied along the *x* and *y* axes to restrict the ligand to an area that is easily determined and capable of being sampled for the length of the simulations. The free energy change of binding ΔG_{PMF} between the bound and unbound sections of the PMF

is given by

$$\Delta G_{\mathrm{PMF}} = -RT \ln\left(\frac{Q_{\mathrm{b,R}}}{Q_{\mathrm{u,R}}}\right) \quad (1)$$

where $Q_{\mathrm{b,R}}$ and $Q_{\mathrm{u,R}}$ are the partition functions for the bound and unbound regions, respectively, and the subscript R denotes the use of orthogonal restraints. Their ratio is computed by integrating the PMF over the bound and unbound regions using the following equation:

$$\frac{Q_{\mathrm{b,R}}}{Q_{\mathrm{u,R}}} = \frac{\int_{\mathrm{bound}} \exp\left(\frac{-W_{\mathrm{R}}(z)}{RT}\right) \mathrm{d}z}{\int_{\mathrm{unbound}} \exp\left(\frac{-W_{\mathrm{R}}(z)}{RT}\right) \mathrm{d}z} = \frac{l_{\mathrm{b}}}{l_{\mathrm{u}}} \exp\left(\frac{-\Delta W_{\mathrm{R}}}{RT}\right) \quad (2)$$

where $W_{\mathrm{R}}(z)$ is the PMF as a function of z and defined to be zero at its lowest point when the ligand is bound. The cutoff between the bound and unbound regions is chosen to be the value of z where the PMF becomes constant within statistical noise and the ligand is in the bulk. The unbound region is defined up to an arbitrary upper limit that cancels in the full expression for ΔG° given below. The PMF depth ΔW_{R} is defined as the lowest point, zero, minus the exponential average of the PMF over the entire unbound region, giving

$$\Delta W_{\mathrm{R}} = RT \ln\left[\int_{\mathrm{unbound}} \exp\left(\frac{-W_{\mathrm{R}}(z)}{RT}\right) \mathrm{d}z / \int_{\mathrm{unbound}} \mathrm{d}z\right] \quad (3)$$

The bound and unbound lengths, l_{b} and l_{u}, are configurational integrals of the PMF

$$l_{\mathrm{b}} = \int_{\mathrm{bound}} \exp\left(\frac{-W_{\mathrm{R}}(z)}{RT}\right) \mathrm{d}z \quad (4)$$

$$l_{\mathrm{u}} = \int_{\mathrm{unbound}} \exp\left(\frac{-(W_{\mathrm{R}}(z) + \Delta W_{\mathrm{R}})}{RT}\right) \mathrm{d}z = \int_{\mathrm{unbound}} \mathrm{d}z \quad (5)$$

which are both independent of the orthogonal restraints. The definition of ΔW_{R} ensures that l_{u} equals the length of the reaction coordinate in the unbound region. The lowest values for $W(z)$ in the binding site contribute the most to the integrals and make the calculations of the ratio of the partition functions insensitive to the cutoff.[42] To determine the standard free energy of binding ΔG°, the free energy ΔG_{V} for changing from the standard-state volume $V^\circ = 1661$ Å^3, which corresponds to a 1 M concentration, to the sampled unbound volume and the free energy ΔG_{R} to remove the orthogonal restraints when the ligand is bound are included to give the expression

$$\Delta G^\circ = \Delta G_{\mathrm{PMF}} + \Delta G_{\mathrm{V}} + \Delta G_{\mathrm{R}} \quad (6)$$

These three free energy terms are depicted in Figure 2. The second term of eq 6, ΔG_{V}, is the ratio of the sampled unbound volume $V_{\mathrm{u,R}}$ to the standard-state volume V° and can be written as

$$\Delta G_{\mathrm{V}} = -RT \ln\left(\frac{V_{\mathrm{u,R}}}{V^\circ}\right) \quad (7)$$

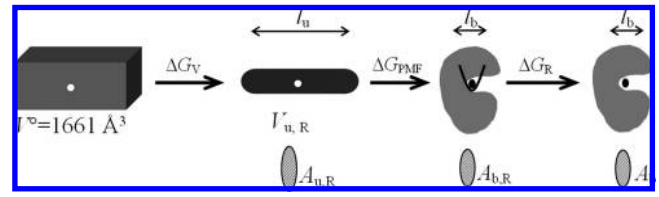

Figure 2. Schematic showing the decomposition of the standard Gibbs free energy of binding ΔG° into the volume reduction term ΔG_{V}, the potential of mean force term ΔG_{PMF}, and the term for the removal of the orthogonal restraints ΔG_{R}.

The unbound volume $V_{\mathrm{u,R}}$ is determined from the distances that the ligand samples along the x, y, and z directions. In the x and y directions, the area $A_{\mathrm{u,R}}$ for the unbound ligand is obtained from the partition function of the orthogonal restraint potential, which is given by

$$A_{\mathrm{u,R}} = \int_{-\infty}^{+\infty} \exp\left(\frac{-k_{xy}x^2}{2RT}\right) \mathrm{d}x \int_{-\infty}^{+\infty} \exp\left(\frac{-k_{xy}y^2}{2RT}\right) \mathrm{d}y = \frac{2\pi RT}{k_{xy}} \quad (8)$$

The reaction coordinate distance spanning the unbound region in the z direction equals l_{u}. $V_{\mathrm{u,R}}$ is simply the distance l_{u} multiplied by the area in the x and y directions

$$V_{\mathrm{u,R}} = l_{\mathrm{u}} \times A_{\mathrm{u,R}} = l_{\mathrm{u}} \frac{2\pi RT}{k_{xy}} \quad (9)$$

ΔG_{R} is calculated using a free energy perturbation approach from the exponential average of turning on the harmonic restraint as follows:

$$\Delta G_{\mathrm{R}} = RT \ln\left\langle \exp\left(\frac{-k_{xy}(\Delta x^2 + \Delta y^2)}{2RT}\right)\right\rangle_{k_{xy}=0} \quad (10)$$

where Δx and Δy are the displacements relative to the restraint minimum for the simulation without orthogonal restraints. The perturbation is carried out using the simulation without restraints as the reference state because the configuration space sampled will include all the important subset of states when the restraint is applied. A perturbation in the reverse direction would not sample all the important states when the restraint is removed. Substituting ΔG_{PMF} and ΔG_{V} from eqs 1, 2, 7, and 9 into eq 6 gives a clear relationship between the standard free energy of binding and the PMF depth as

$$\Delta G^\circ = \Delta W_{\mathrm{R}} - RT \ln\left(\frac{l_{\mathrm{b}} A_{\mathrm{u,R}}}{V^\circ}\right) + \Delta G_{\mathrm{R}} \quad (11)$$

which is independent of length l_{u}. The equivalent binding equilibrium constant, K_{a}, is

$$K_{\mathrm{a}} = \frac{l_{\mathrm{b}} A_{\mathrm{u,R}}}{V^\circ} \exp\left(\frac{\Delta G_{\mathrm{R}} - \Delta W_{\mathrm{R}}}{RT}\right) \quad (12)$$

Equation 12 has the following interpretation: ΔW_{R} does not contain the integration of the PMF over the bound length so this must be included in the form of l_{b}; it does contain

integration over $A_{u,R}$ so this must be removed by multiplication by $A_{u,R}$ (division by $1/A_{u,R}$) since it concerns the unbound state; it does not contain integration over V° when unbound so this must be included by dividing by V°; it contains integration over the orthogonal restraints when bound so this contribution is removed by multiplying by $\exp(\Delta G_R/RT)$. Defining ΔG_R in terms of the ratio of orthogonal areas, we have

$$\Delta G_R = -RT \ln\left(\frac{A_{b,R}}{A_b}\right) \tag{13}$$

where $A_{b,R}$ and A_b are the bound areas available to the ligand with and without the orthogonal restraints, and setting $V_b = l_b A_b$ leads to the insightful relationship

$$\Delta G^\circ = \Delta W_R - RT \ln\left(\frac{V_b}{V^\circ}\frac{A_{u,R}}{A_{b,R}}\right) \tag{14}$$

If the PMF is three-dimensional in x, y, and z, then an analogous expression can be derived for ΔG° as

$$\Delta G^\circ = \Delta W_{3D} - RT \ln\left(\frac{V_b}{V^\circ}\right) \tag{15}$$

where ΔW_{3D} is the minimum in the three-dimensional PMF obtained by integrating over all system coordinates except x, y, and z of the ligand. Equating eqs 14 and 15 yields the relationship

$$\Delta W_{3D} = \Delta W_R - RT \ln\left(\frac{A_{u,R}}{A_{b,R}}\right) \tag{16}$$

Model Binding Site. The approach is first applied to a model binding site. The three-dimensional PMF, as shown in Figure 1c, is harmonic in the x and y directions with a force constant k_b of 20 kcal mol^{-1} Å^{-2} and a step function in z with a length l_b of 0.5 Å. ΔW_{3D} is chosen to be -10 kcal mol^{-1}. Equation 16 is used to calculate the value of ΔW_R. Equation 8 gives the required areas $A_{u,R}$ and $A_{b,R}$ and the force constant when bound equals $k_{xy} + k_b$. Five orthogonal force constants k_{xy} are tested: 0, 1, 5, 10, and 50 kcal mol^{-1} Å^{-2}.

Trypsin−Benzamidine Setup and Molecular Dynamics Protocol. The protein−ligand system examined is the inhibitor benzamidine binding to the enzyme bovine trypsin (PDB code 3PTB) whose function is to degrade dietary proteins. This system has commonly been used as a benchmark for computational methods.[33,43−46] The same five orthogonal force constants k_{xy} are tested: 0, 1, 5, 10, and 50 kcal mol^{-1} Å^{-2}. The protein is modeled with the AMBER 99 force field[47] and solvated with 9284 TIP3P[48] water molecules using the leap module of AMBER 9.[49] Nine Cl^- ions are added to neutralize the system. The ligand, benzamidine, is modeled with the GAFF force field,[50] and its charges are derived using the RESP method.[51] This setup results in a system comprising 31 100 atoms in a box of size $66 \times 62 \times 75$ Å^3.

The energy minimization and molecular dynamics simulations are performed with the Sander module of AMBER 9. Periodic boundary conditions and the particle mesh Ewald method are used with a nonbonded cutoff of 9 Å. The system is first energy-minimized using 1000 steps of each of the steepest descent and conjugate gradient methods. It is then heated at constant volume to 298 K in a 50-ps molecular dynamics simulation with Langevin temperature regulation. Bonds involving hydrogen atoms are constrained using the SHAKE algorithm allowing a 2-fs integration time step. Subsequently, the system is switched to a constant pressure of 1 bar and further equilibrated for 1 ns.

Umbrella Sampling. Umbrella sampling is adopted to determine the PMF from the binding site to solution in water, whereby a harmonic restraint is placed at successive points along the reaction coordinate ξ and probabilities at each point are accumulated. The restraining potential employed has the harmonic form

$$V(\xi) = \frac{1}{2}k(\xi - \xi_0)^2 \tag{17}$$

where ξ_0 is the target distance and k is the force constant. The weighted histogram analysis method[52,53] is used to convert the probabilities into the PMF along the reaction coordinate. The reaction coordinate here is defined as either the z-component of the distance between the C7 atom of benzamidine and the Cγ atom of ASP 189, as shown in Figure 1d, or the radial distance r. The radial distance when the ligand is bound is ~3.8 Å. The path extends linearly out of the binding site and is divided into 28 windows at 1 Å intervals, with each window having the ligand placed at the desired value of the reaction coordinate, starting from 4 to 31 Å, remaining well within the simulation box. The restraint force constant of the umbrella potential is set to 1 kcal mol^{-1} Å^{-2}. Each window is simulated for 1 ns.

Choice of the Binding Path. A suitable linear path, to which this approach is applied, was chosen by performing short umbrella sampling simulations along 17 different linear paths at different orientations from the binding pocket to the bulk as illustrated in Figure 3. One of these trajectories is along the central z-axis, which coincides with the long axis of the ligand when bound. The other trajectories are chosen with combinations of four different polar angles θ (5°, 10°, 15°, and 20°) and four different azimuthal angles φ (0°, 90°, 180°, and 270°). The radial distance reaction coordinate is employed. Each window on every path is simulated for only 50 ps to offset the expense of simulating the larger number of paths. An alternative approach where each window is started from the last configuration of the preceding window to create a nonlinear path was also examined but was found to be dissatisfactory, as reported in the results.

Results

Standard Free Energy of Binding of the Model System. Figure 4 shows the PMFs for the model system with four different orthogonal restraints. The PMF depths ΔW_R are seen to decrease with smaller orthogonal force constant k_{xy} and tend to negative infinity in the limit of $k_{xy} = 0$ when the unbound ligand samples an infinite volume. Using the arbitrary unbound length l_u of 10 Å, Table 1 summarizes l_b,

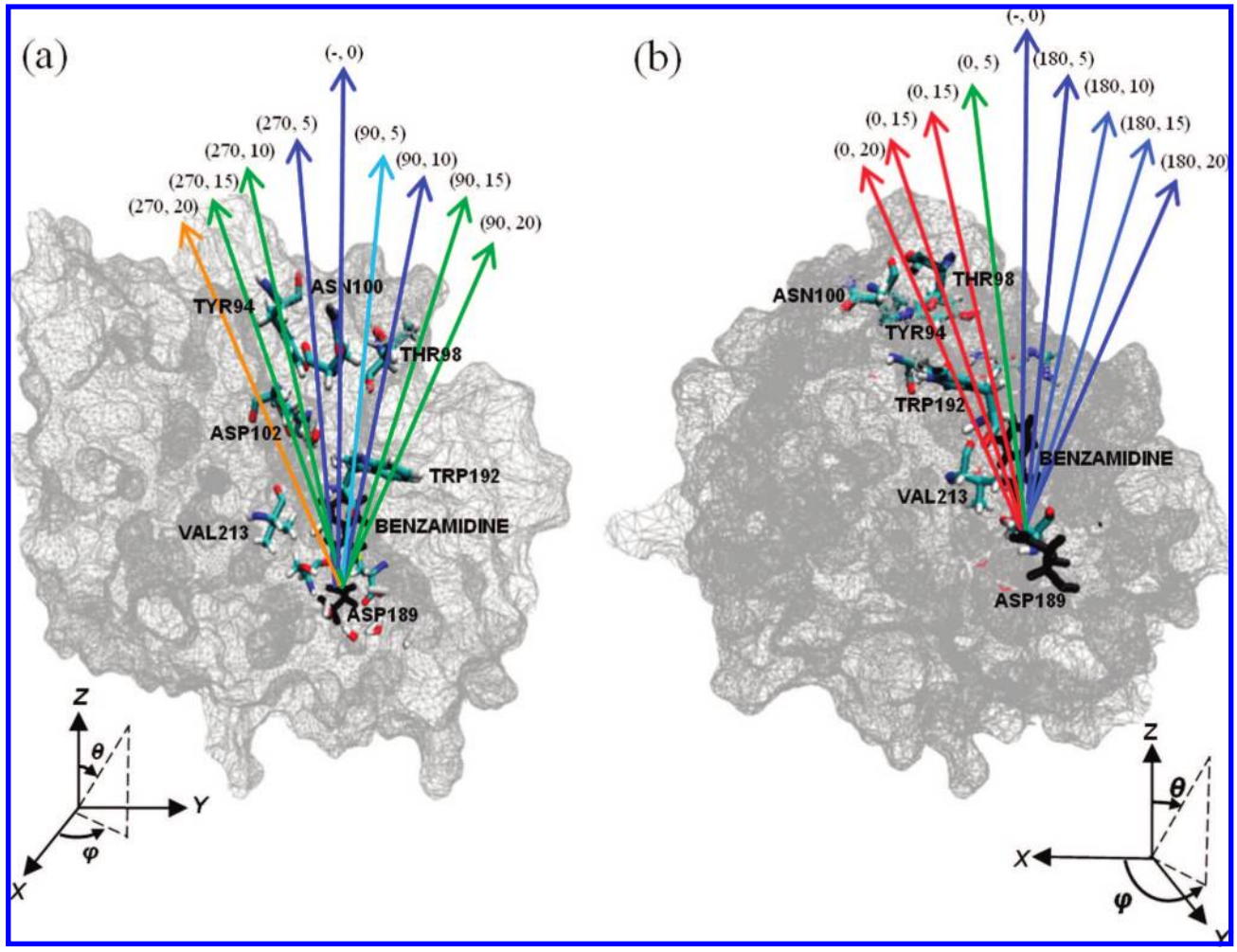

Figure 3. Seventeen paths considered for umbrella sampling that connect the binding site to the bulk in the (a) *yz* plane and (b) *xz* plane. Each path is labeled by its azimuthal and polar angles relative to the central path. Asp 189 and benzamidine are in black. Paths are colored according to their PMF depths, ΔW_R, with blue the deepest and red the shallowest.

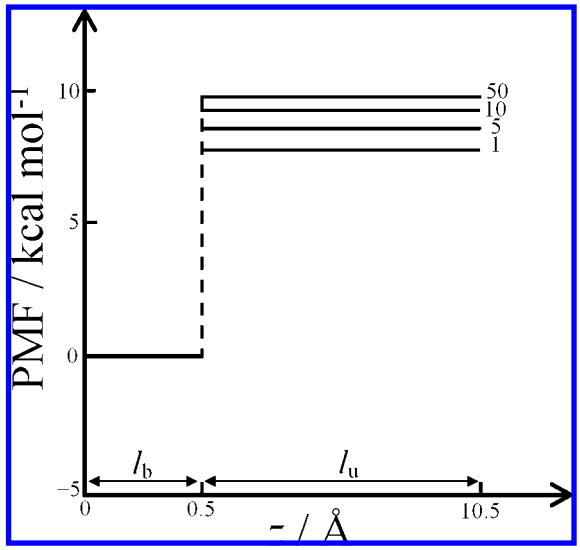

Figure 4. PMFs along the *z*-component reaction coordinate for the model binding system with orthogonal force constants of 1, 5, 10, and 50 kcal mol^{-1} Å^{-2}.

$A_{u,R}$, ΔW_R, ΔG_{PMF}, ΔG_V, ΔG_R, and $\Delta G°$ for different values of k_{xy}. The unbound area $A_{u,R}$ orthogonal to the reaction coordinate decreases with larger k_{xy}, and ΔW_R and ΔG_{PMF} become more favorable as would be expected since the ligand is binding from a smaller unbound volume. The unbound volumes $V_{u,R}$ range from 37 Å^3 for k_{xy} = 1 kcal mol^{-1} Å^{-2} to 0.74 Å^3 for k_{xy} = 50 kcal mol^{-1} Å^{-2}. Consequently, the free energy change ΔG_V for contracting from the standard-state volume to the unbound volume increases from 2.2 to 4.5 kcal mol^{-1}. The free energy change ΔG_R for turning off the restraints in the binding site decreases from 0.0 to −0.7 kcal mol^{-1} with larger k_{xy}. The same value is obtained for the standard free energy of binding $\Delta G° = -4.2$ kcal mol^{-1}, regardless of k_{xy}. This value is consistent with that predicted by eq 15 using $V_b = 0.092$ Å^3 as determined by an analogue of eq 9. It would also be independent of l_u because this cancels between ΔG_{PMF} and ΔG_V.

Standard Free Energy of Binding of Trypsin−Benzamidine. Figure 5a−e shows the PMFs obtained using a *z*-component reaction coordinate with orthogonal restraints of 0, 1, 5, 10, and 50 kcal mol^{-1} Å^{-2}, respectively, for different sampling times for the trypsin−benzamidine system. The central path was used for this run since it is one of the most favorable paths. If the other five most favorable paths are extended to 1-ns sampling per window, similar ΔW_R results are found, confirming the choice of the central path. A more detailed examination of the alternative paths is

Table 1. Standard Free Energy of Binding ΔG° and Its Components for Different Orthogonal Restraints k_{xy} for the Model System

k_{xy}/(kcal mol^{-1} Å^{-2})	l_b/(Å)	$A_{u,R}$/(Å^2)	ΔW_R/(kcal mol^{-1})	ΔG_{PMF}/(kcal mol^{-1})	ΔG_v/(kcal mol^{-1})	ΔG_R/(kcal mol^{-1})	ΔG°/(kcal mol^{-1})
0	0.5	$+\infty$	$+\infty$	$+\infty$	$-\infty$	0.0	
1	0.5	3.7	−8.2	−6.4	2.2	−0.0	−4.2
5	0.5	0.74	−9.0	−7.3	3.2	−0.1	−4.2
10	0.5	0.37	−9.3	−7.6	3.6	−0.2	−4.2
50	0.5	0.074	−9.8	−8.0	4.5	−0.7	−4.2

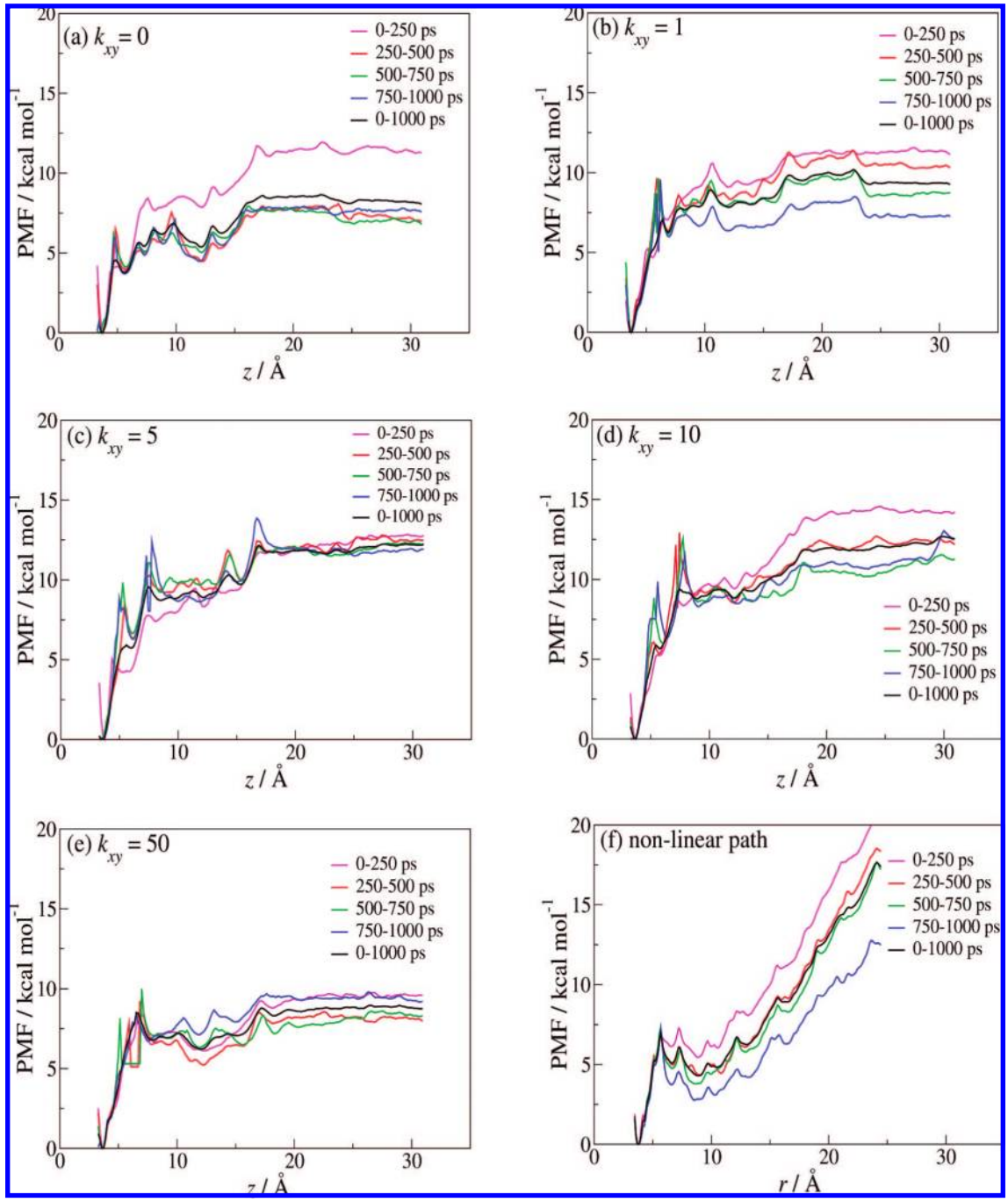

Figure 5. PMFs from umbrella sampling for trypsin−benzamidine using the *z*-component reaction coordinate with an orthogonal force constant k_{xy} (kcal mol^{-1} Å^{-2}) of (a) 0, (b) 1, (c) 5, (d) 10, and (e) 50. (f) PMF using the radial reaction coordinate and the nonlinear path, for which each window starts from the last frame of the preceding one. PMFs are shown averaged over 250-ps intervals (magenta, red, green, blue) and over the full 1 ns (black).

provided below. The PMFs with orthogonal restraints of 5 and 50 kcal mol^{-1} Å^{-2} are seen to converge slightly better than those at 0, 1, and 10 kcal mol^{-1} Å^{-2}. We might expect that the convergence would be worse for weaker restraints since the ligand has to sample a greater volume. In the limit of zero force constant, this volume would be infinite.

Table 2. Standard Free Energy of Binding $\Delta G°$ and Its Components for Different Orthogonal Restraints k_{xy} for the Trypsin−Benzamidine System

k_{xy}/(kcal mol^{-1} Å^{-2})	l_b/(Å)	$A_{u,R}$/(Å^2)	ΔW_R/(kcal mol^{-1})	ΔG_{PMF}/(kcal mol^{-1})	ΔG_V/(kcal mol^{-1})	ΔG_R/(kcal mol^{-1})	$\Delta G°$/(kcal mol^{-1})
0	0.54		−8.3 (1.0)	−6.4			
1	0.41	3.7	−9.6 (0.8)	−7.5	2.0	−1.2 (0.4)	−6.7 (0.9)
5	0.59	0.74	−11.9 (0.1)	−10.0	3.0	−2.0 (0.5)	−9.0 (0.5)
10	0.53	0.37	−11.9 (0.8)	−10.0	3.4	−2.3 (0.6)	−8.9 (1.0)
50	0.49	0.074	−8.7 (0.4)	−6.7	4.4	−3.2 (0.6)	−5.5 (0.7)

However, for the 1-ns trajectories run here, a significant difference is not observed.

Table 2 summarizes l_b, $A_{u,R}$, ΔW_R, ΔG_{PMF}, ΔG_V, ΔG_R, and $\Delta G°$ for different values of k_{xy}. The cutoff between the bound and unbound regions of the PMF is chosen to be 17 Å and the maximum value of the reaction coordinate is chosen to be 31 Å, so the length l_u in the unbound region equals 14 Å. The bound length l_b is similar for different orthogonal restraints. The unbound area $A_{u,R}$ is unchanged from the model system. ΔW_R and ΔG_{PMF} again become more favorable with larger k_{xy}, with the exception of a less favorable value of ΔW_R for $k_{xy} = 50$ kcal mol^{-1} Å^{-2}. ΔG_V increases from 2.0 to 4.4 kcal mol^{-1}, while ΔG_R decreases from −1.2 to −3.2 kcal mol^{-1} for the same range of k_{xy}. These values of ΔG_R are larger than those in the model system, presumably because the minimum of the orthogonal restraint in the trypsin−benzamidine system no longer matches up with the effective minimum of the binding site as it was assumed to do in the model system. The final standard free energies of binding $\Delta G°$ range from −5.5 to −9.0 kcal mol^{-1}, comparing reasonably well with the experimental value[54,55] of −6.3 kcal mol^{-1}. Our estimates of the standard errors of the mean free energy differences ΔW_R are given in parentheses in Table 2 and utilize the four 250-ps time frames (Figure 5) to estimate errors due to inadequate sampling. The errors on l_b are estimated in the same manner and are found to be negligible compared to ΔW_R; standard errors in ΔG_R are also taken over the 250-ps time periods. The final errors on $\Delta G°$ are found to lie between 0.5 and 1.0 kcal mol^{-1}. Other errors arising from the force field accuracy and possible incomplete sampling cannot be reliably estimated.

Other Linear and Nonlinear Paths. The PMF along the nonlinear radial path for trypsin−benzamidine is displayed in Figure 5f, averaged over different time slices of the simulations. There are two clear problems with this PMF: the first is that it is not converged and decreases with time; the second is that the PMF does not become constant at long range as would be expected once the ligand is fully solvated. This is because the ligand never actually leaves the protein surface over the entire 25 Å reaction coordinate. Although such paths are possibly more realistic in describing binding, such an approach was not pursued further due to the excessively long reaction coordinate, poor convergence, and difficulty in defining the unbound volume.

For the linear paths approach, the PMF values of all 17 paths are color-coded and displayed in Figure 6. In principle, all PMFs should produce the same ΔW_R; however, in practice, they do not due to convergence problems arising from the short simulation timescales and steric clashes between the ligand and protein. The PMFs of all paths are plotted in Figure 6 at 1 Å intervals with a common reference point in the bulk solvent. The lowest values of the PMFs are colored blue and the highest are colored red. The plot shows that paths with $\varphi = 180°$ all have favorable PMFs with large binding energies, and Figure 3b indicates that this

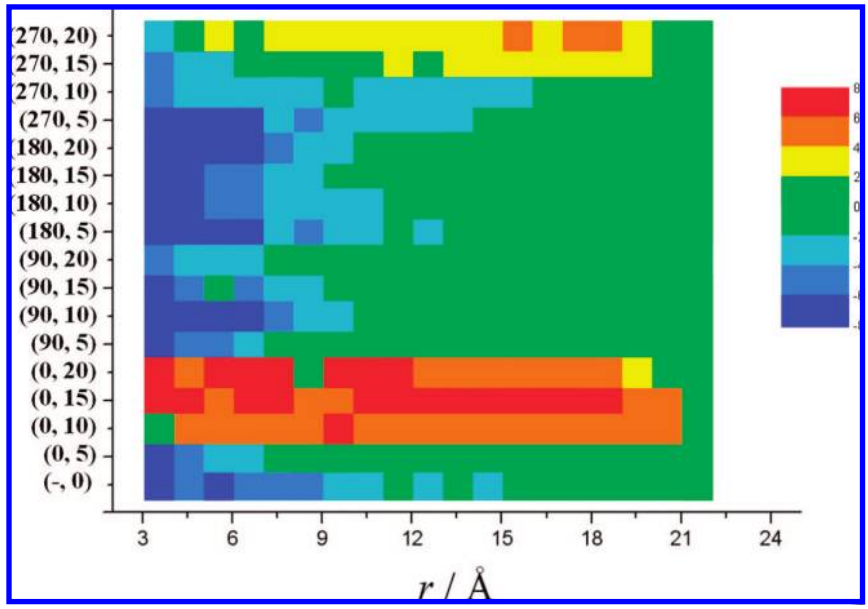

Figure 6. Color-coded PMFs along the distance reaction coordinate for each of the 17 linear paths in the (φ, θ) notation. The points with the blue color have the lowest value of $W(z)$ and those with red the highest.

is the open end of the protein where fewer residues occupy the path of the ligand. Other favorable paths include (−, 0°), (90°, 10°), (180°, 5°), and (270°, 5°). Most paths give conventional binding PMFs in that there is a minimum at the bound state and the PMF curve flattens off in the unbound region. However, for paths having $\varphi = 0°$ and $\theta = 10°$, 15°, or 20°, a tryptophan residue (Trp 192 in Figure 3) blocks the way of the ligand. This forces it away from these paths, resulting in unreasonable PMFs for the short timescale of simulation used.

Discussion

In this work, we show how to obtain the standard free energy of binding $\Delta G°$ from a one-dimensional PMF and apply it to a model binding system and the trypsin−benzamidine system. The PMF along the *z*-component of the reaction coordinate is calculated by applying restraints to the ligand orthogonal to the reaction coordinate. We calculate the ratio of the partition functions of the bound and unbound states from the PMF. The sampled unbound volume is computed from the volume spanned within the orthogonal restraint along the unbound section of the PMF. $\Delta G°$ is obtained from the free energy of contraction from the standard volume to the unbound volume, then to the bound state with restraints, and then to the bound state. Our approach gives reasonable results in comparison with experiment for the trypsin−benzamidine system,[54,55] and the results are reasonably independent of the size of the orthogonal restraint. The main exception appears to be for $k_{xy} = 50$ kcal mol^{-1} Å^{-2} for which a less favorable value of $\Delta G°$ is obtained. This may arise because the larger force constant is constraining the ligand too tightly to a linear path, bringing it too close to the protein, and leading to a nonconverged PMF.

It is clear that $\Delta G°$ is not ΔW_R as is often assumed.[31−34] Nor is it always obtained by augmenting ΔW_R with the change in free energy from the standard volume to the bound volume. Because the one-dimensional PMF includes averaging over the ligand's orthogonal coordinates *x* and *y*, one must know the *xy* area sampled in the unbound region to work out the volume sampled in the unbound region and therefore determine the free energy for scaling to it from the standard volume. Equation 14 for $\Delta G°$ becomes equivalent to that of Lee and Olson[12,18] when $A_{u,R} = A_{b,R}$. This equality will hold in the limit of an infinitely large orthogonal force constant, when confinement by the protein would be negligible, or when sampling is very short. Usually $A_{u,R}$ would exceed $A_{b,R}$ because the binding site also confines the ligand, and so omitting this term would overestimate $\Delta G°$ and make it less favorable. Short sampling may not allow the ligand to adequately explore the larger $A_{u,R}$. It is curious to note that the $\Delta G°$ values obtained do not differ substantially from ΔW_R obtained without an orthogonal restraint or any consideration of the standard-state concentration. This is purely coincidental. If a different standard-state concentration were used, $\Delta G°$ would change but ΔW_R would not. Use of a three-dimensional PMF and eq 15 to get $\Delta G°$ would avoid the need to define an orthogonal restraint. However, this suffers from the disadvantage in the case of umbrella sampling that many more umbrella potentials would be required to span the three-dimensional space. Similar issues apply to the rotational entropy. If an angular restraint were to be applied to sample smaller angular volumes, the free energy change from $8\pi^2$ in solution to the unbound angular volume would be required. The change in rotational entropy here is assumed to be included in ΔW_R.

Comparing the approach of Woo and Roux,[21] their derivation includes the calculation of a surface area integral at the unbound state and a PMF integral over the bound region. They apply translational, rotational, and conformational restraints on the large ligand used in that study. Their partition function ratio is made up of four terms: S^*, a surface area integral over what is effectively a patch of a sphere of arbitrary radius 40 Å, and I^*, the integral of the PMF over the bound region up to this radius, and free energy terms for turning on and off the restraints. S^* corresponds to our $A_{u,R}$ and I^* to our $l_b \exp(-\Delta W_R/RT)$. Our method is a simpler formulation expressed in terms that have a clear physical interpretation. The use of a *z*-component versus a radial reaction coordinate leads to a number of advantages: there is a simple expression for the unbound area $A_{u,R}$ independent of the cutoff; there is no need for a Jacobian, which results in a flat PMF in the unbound region; the area to be sampled is roughly constant at each value of the reaction coordinate; and ΔW_R is exponentially averaged over the entire unbound region, reducing its statistical error. The interdependence of I^* and S^* on the cutoff and the dependence of the PMF on the radial Jacobian make it difficult to consider more than one radial point in the unbound region. Combining these attributes of the method gives a simple relationship between $\Delta G°$ and ΔW_R (eq 11 or 14).

The trypsin−benzamidine system has been previously studied using umbrella sampling and grand canonical simulations by Resat et al.[45] They obtained a rather different PMF with a depth of −25 kcal mol^{-1} and a pronounced barrier of 10 kcal mol^{-1} at a reaction coordinate distance of ∼5 Å. The larger PMF depth could be due to less equilibration and a rigid protein. Gervasio et al.[33] studied the same system with metadynamics using a two-dimensional reaction coordinate made up of a distance and a ligand orientation. They found good agreement with experiment although it should be noted that they do not include the standard-state dependence in their comparison.

We also examined multiple unbinding pathways of benzamidine from trypsin. One protocol that involves no assumption of the unbinding pathway uses a nonlinear radial trajectory, similar to the approach of Hajjar et al.[56] Having a nonlinear umbrella-sampling trajectory would give more realistic information on the binding mechanism; however, this approach has a number of shortcomings: it results in deep, unconverged PMFs and would require a long reaction coordinate to the bulk since the ligand remains bound to the protein surface even at large distances. Serial simulations are also required since each window is started from the preceding one. Despite all the disadvantages of this protocol, it may more accurately represent the binding mechanism whereby the ligand may first bind to the protein somewhere on the surface at a distance from the binding site rather then directly into the binding site from solution. Calculating the

PMF of the ligand along linear pathways avoids the problem in the nonlinear protocol because the ligand reaches the bulk solvent along a much shorter reaction coordinate and each window, being independent of the previous one, can be run in parallel. Multiple trajectories were also used by Karplus and van der Vaart[57] but to obtain a PMF in targeted MD calculations. Hajjar et al.[56] use multiple trajectory umbrella sampling in addition to nonlinear paths. The linear paths method allows us to identify the most favorable binding paths that the ligand can follow. It also generates an approximate free energy landscape of the ligand near the protein's binding site. This landscape tells us the expected result that PMF is governed by sterics so that the lowest and most converged PMFs span regions that are least occupied by the protein. One path is then chosen to estimate the standard free energy of binding using the z-component reaction coordinate and orthogonal restraints as discussed earlier. Although the PMFs may appear barrierless, they represent an average over many configurations and do not necessarily imply that individual binding or unbinding trajectories have no barriers.

Conclusions

We introduce an approach to calculate the standard free energy of binding $\Delta G°$ from a one-dimensional potential of mean force. The PMF is calculated using a z-component reaction coordinate with orthogonal restraints. The ratio of the bound and unbound partition functions is calculated from the PMF, and the standard-state concentration is accounted for by determining the unbound volume sampled by the ligand through application of restraining potentials. The final expression for $\Delta G°$ depends on the bound length along the reaction coordinate, the unbound area orthogonal to the reaction coordinate, the standard volume, the orthogonal restraining potential, and the PMF depth ΔW_R. We demonstrate how to get $\Delta G°$ from the PMF for a model binding system and go on to apply it to the PMF of benzamidine binding to trypsin to predict $\Delta G°$ to range from -5.5 to -9.0 kcal mol^{-1}, in good agreement with experiment at -6.3 kcal mol^{-1}. We emphasize how the formulation is simpler when using a z-component reaction coordinate rather than a radial one and make clear how the formulation depends on the dimensionality of the PMF.

Acknowledgment. S.D. thanks the School of Chemistry at the University of Manchester, the Algerian government for funding, and also Dr. Richard Dimelow for providing the WHAM code to generate the PMFs from the umbrella sampling simulations. R.H.H. is funded by EPSRC Grant EP/E026222/1.

References

(1) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. *Adv. Protein Chem.* **2003**, *66*, 123–158.

(2) Chipot, C.; Pohorille, A. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer: Berlin, 2007.

(3) Gilson, M. K.; Zhou, H.-X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.

(4) Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644–2676.

(5) Meirovitch, H. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181–186.

(6) Raha, K.; Merz, J. K. M.; David, C. S. *Annual Reports in Computational Chemistry*; Elsevier: Amsterdam, 2005; pp 113−130.

(7) Rodinger, T.; Pomès, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 164–170.

(8) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. *Annual Reports in Computational Chemistry*; Elsevier: Amsterdam, 2007; pp 41−59.

(9) Wong, C. F.; McCammon, J. A. *Adv. Protein Chem.* **2003**, *66*, 87–121.

(10) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.

(11) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67–74.

(12) Lee, M. S.; Olson, M. A. *Biophys. J.* **2006**, *90*, 864–877.

(13) Åqvist, J.; Medina, C.; Samuelsson, J.-E. *Protein Eng.* **1994**, *7*, 385–391.

(14) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742–3746.

(15) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798–2814.

(16) Bui, J. M.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2003**, *85*, 2267–2272.

(17) Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 503–507.

(18) Lee, M. S.; Olson, M. A. *J. Phys. Chem. B* **2008**, *112*, 13411–13417.

(19) Palma, R.; Himmel, M. E.; Brady, J. W. *J. Phys. Chem. B* **2000**, *104*, 7228–7234.

(20) Tsunekawa, N.; Miyagawa, H.; Kitamura, K.; Hiwatari, Y. *J. Chem. Phys.* **2002**, *116*, 6725–6730.

(21) Woo, H. J.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825–6830.

(22) Bolhuis, P. G.; Dellago, C.; Chandler, D. *Faraday Discuss.* **1998**, *110*, 421–436.

(23) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578–581.

(24) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169–9183.

(25) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.

(26) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

(27) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.

(28) Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.

(29) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.

(30) Pranata, J.; Jorgensen, W. L. *Tetrahedron* **1991**, *47*, 2491–2501.

(31) Baştuğ, T.; Chen, P.-C.; Patra, S. M.; Kuyucak, S. *J. Chem. Phys.* **2008**, *128*, 155104.

(32) Charlier, L.; Nespoulous, C.; Fiorucci, S.; Antonczak, S.; Golebiowski, J. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5761–5771.

(33) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.

(34) Ghoufi, A.; Malfreyt, P. *J. Chem. Phys.* **2006**, *125*, 224503.

(35) Justice, M. C.; Justice, J. C. *J. Solution Chem.* **1976**, *5*, 543–561.

(36) Prue, J. E. *J. Chem. Educ.* **1969**, *46*, 12–16.

(37) Shoup, D.; Szabo, A. *Biophys. J.* **1982**, *40*, 33–39.

(38) Khavrutskii, I. V.; Dzubiella, J.; McCammon, J. A. *J. Chem. Phys.* **2008**, *128*, 044106.

(39) Trzesniak, D.; Kunz, A. P. E.; van Gunsteren, W. F. *Chem. Phys. Chem.* **2007**, *8*, 162–169.

(40) Dang, L. X.; Pettitt, B. M. *J. Am. Chem. Soc.* **1987**, *109*, 5531–5532.

(41) Setny, P. *J. Chem. Phys.* **2008**, *128*, 125105.

(42) Mihailescu, M.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 23–36.

(43) Essex, J. W.; Severance, D. L.; TiradoRives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **1997**, *101*, 9663–9669.

(44) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6290–6295.

(45) Resat, H.; Marrone, T. J.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 522–532.

(46) Schwarzl, S. M.; Tschopp, T. B.; Smith, J. C.; Fischer, S. *J. Comput. Chem.* **2002**, *23*, 1143–1149.

(47) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

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

(49) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. AMBER9; University of California, San Francisco: San Francisco, CA, 2006.

(50) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

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

(52) Dimelow, R. J.; Bryce, R. A.; Masters, A. J.; Hillier, I. H.; Burton, N. A. *J. Chem. Phys.* **2006**, *124*, 114113.

(53) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(54) Mares-Guia, M. *Arch. Biochem. Biophys.* **1968**, *127*, 317–322.

(55) Talhout, R.; Engberts, J. B. F. N. *Eur. J. Biochem.* **2001**, *268*, 1554–1560.

(56) Hajjar, E.; Perahia, D.; Debat, H.; Nespoulous, C.; Robert, C. H. *J. Biol. Chem.* **2006**, *281*, 29929–29937.

(57) van der Vaart, A.; Karplus, M. *J. Chem. Phys.* **2007**, *126*, 164106.

CT8002354