Absolute Binding Free Energies: A Quantitative Approach for Their Calculation

Stefan Boresch,* Franz Tettinger, and Martin Leitgeb

Molecular Dynamics and Biomolecular Simulation Group, Department of Theoretical Chemistry and Molecular Structural Biology, University of Vienna, Währingerstrasse 17, 1090 Vienna, Austria

Martin Karplus*

Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France, and Department of Chemical Biology, 12 Oxford Street, Harvard University, Cambridge, Massachusetts 02138

Received: August 1, 2002; In Final Form: February 21, 2003

The computation of absolute binding affinities by molecular dynamics (MD) based free energy simulations is analyzed, and an exact method to carry out such a computation is presented. The key to obtaining converged results is the introduction of suitable, auxiliary restraints to prevent the ligand from leaving the binding site when the native ligand−receptor interactions are turned off alchemically. We describe a versatile set of restraints that (i) can be used in MD simulations, that (ii) restricts both the position and the orientation of the ligand, and that (iii) is defined relative to the receptor rather than relative to a fixed point in space. The free energy cost, ΔA_r, for this set of restraints can be evaluated analytically. Although the techniques were originally developed for the gas phase, the resulting expression is exact, since all contributions from solute−solvent interactions cancel from the final result. The value of ΔA_r depends only on the equilibrium values and force constants of the chosen harmonic restraint terms and, therefore, can be easily calculated. The standard state dependence of binding free energies is also investigated, and it is shown that the present approach takes this into account correctly. The analytical expression for ΔA_r is verified numerically by calculations on the complex formed by benzene with the L99A mutant of T4 lysozyme. The overall approach is illustrated by a complete binding free energy calculation for a complex based on a simplified model for tyrosine bound to tyrosyl-tRNA-synthetase. The results demonstrate the usefulness of the proposed set of restraints and confirm that the calculated binding free energy is independent of the details of the restraints. Comparisons are made with earlier formulations for the calculation of binding free energies, and certain limitations of that work are described. The relationship between ΔA_r and the loss of translational and rotational entropy during a binding process is analyzed.

1. Introduction

Free energy difference calculations have become a valuable tool in the arsenal of methods available to the computational chemist and computational molecular biologist.[1−4] Successful applications to many diverse areas have been reported, including the study of solvation free energies,[5,6] the thermodynamics of ligand binding,[7−9] the effects of point mutations in proteins,[9] and conformational equilibria.[10] In the vast majority of free energy simulations, double free energy differences $\Delta\Delta A$ were calculated exploiting thermodynamic cycles.[11] A particularly promising aspect of free energy computations is their capability to give not only the (double) binding free energy difference $\Delta\Delta A_{bind}$ for the affinity of two ligands L1 and L2 (to, for example, a protein P) but also the free energy of binding for each of the ligands to P individually; that is, to compute the binding free energy ΔA_{bind} for the process

$$(L)_{H_2O} + (P)_{H_2O} \xrightarrow{\Delta A_{bind}} (PL)_{H_2O} \quad (1)$$

where PL denotes the complex and $(\,)_{H_2O}$ makes clear that the process takes place in aqueous solution. An example where knowing ΔA_{bind} is important arises in F_1ATPase.[12] Free energy calculations, interestingly, provide the only way at present of making the connection between the macroscopic measurements of the experimental binding affinities[12] and the microscopic structural data.[13] Although free energy difference values (e.g. $\Delta\Delta A$ between ATP and ADP + P_i) for binding to the various catalytic subunits have resolved this problem,[14] the absolute binding affinities will provide more detailed information necessary for a full description of the mechanism of this enzyme (Yang et al., work in progress).

The direct calculation of ΔA_{bind} according to eq 1 would require a simulation that starts with P and L bound and then follows the (un)binding process to the completely separated ligands. Because of the complexity of this problem, such a calculation is very difficult computationally, although some attempts in this direction have been published.[15] As for the calculation of relative binding free energy differences,[11] a suitably chosen thermodynamic cycle permits one to overcome this problem. This possibility was first described by Jorgensen et al.[16] The calculation of ΔA_{bind} is split into two parts

$$(L)_{H_2O} \xrightarrow{\Delta A_I} (L)_{gas} \quad (2a)$$

$$(PL)_{H_2O} \xrightarrow{\Delta A_{II}} (P)_{H_2O} + (L)_{gas} \quad (2b)$$

In eq 2a for ΔA_I, the ligand is transferred from the solution

* To whom correspondence should be addressed. E-mail: marci@tammy.harvard.edu. E-mail: stefan@mdy.univie.ac.at. Web site: http://www.mdy.univie.ac.at/en/sbhome.html.

10.1021/jp0217839 CCC: $25.00
Published on Web 08/13/2003

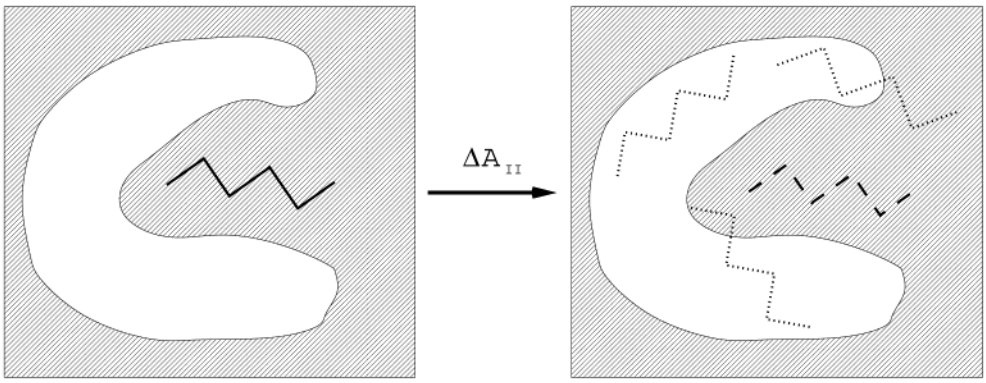

Figure 1. Graphical representation of step II of the DAM thermodynamic cycle (eq 2). Solvent is indicated by the gray background, the protein is shown in white, and the ligand is shown as the thick black line. The left side depicts the native bound state where the protein−ligand interactions keep the ligand in its binding pocket. The right side illustrates the end point of eq 2b, where the protein−ligand interactions have been turned off alchemically. The ligand is free to wander through the simulation system, as indicated by the thinner, dashed copies of the ligand. See the main text for a detailed discussion.

into the gas phase. In the second step, eq 2b for ΔA_{II}, L is initially bound to P and the complex is in solution and, similarly to the first step, L is transferred into the gas phase; that is, it ceases to interact with both P (which remains solvated) and the solvent. Frequently, this transfer into the gas phase is referred to as annihilation, which has led to the name *double annihilation method* (DAM). One sees immediately that the difference of the two steps described in eq 2 gives the desired free energy ΔA_{bind} of the ligand L binding to the protein P with both in solution;[16] that is,

$$\Delta A_{I} - \Delta A_{II} = \Delta A_{bind} \tag{3}$$

A number of applications of the thermodynamic cycle, eq 2, have been reported; see, for example, refs 16−25. However, the direct application of the DAM, as described by eq 2, suffers from two problems.[26−28] First, the free energy of any reaction of the type A + B → AB depends on the standard state;[29] that is, one should more correctly write (and compute) ΔA_{bind}^{0} instead of ΔA_{bind}. The original DAM does not take this standard state dependence into account, and it is not immediately clear where this standard state dependence fits into the DAM framework. The second problem arises during the calculation of ΔA_{II} (eq 2b). To see why this is the case, consider Figure 1, which is a cartoon of the initial (left picture) and final state (right picture) of a free energy simulation attempting to compute ΔA_{II}. At the beginning of the calculation the protein (shown in white) and the ligand (solid black line) interact normally; (assuming reasonably tight binding) the interactions in the physical potential function keep the ligand in the binding pocket. However, at the end point of step II, all interactions between P and L have been turned off alchemically; in the figure this is indicated by using a broken line for L. Thus, the ligand is free to wander to any point in the simulation system, as shown by the grayish replicas of the original ligand (dark, broken line). To compute ΔA_{II} correctly, the ligand would have to sample *every* possible position and orientation relative to the protein. (This is not the case in step I, since the solvent surrounding L is an isotropic medium, so all positions and orientations of L in the simulation system are equivalent.) Such extensive sampling is not feasible computationally; furthermore, when calculating the free energy of the reverse process, one faces the same problem as that for a direct calculation of ΔA_{bind} according to eq 1; that is, the ligand, which at the beginning of the backward calculation is free to move to any point of the simulation system, would have to find a path leading it from an arbitrary starting point to the native binding mode. Indeed, several DAM calculations based on eq 2 either reported no backward calculations, or there was a large hysteresis between forward and backward results; see, for example, refs 17 and 18.

Various methods to circumvent this problem can be found in the literature, and the "wandering" of the ligand at the end point was prevented by different means;[19−22,24,25] for example, Miyamoto and Kollman added five restraints, modeled as hydrogen bonds formed by the ureido group of biotin in the native biotin−streptavidin complex during the "annihilation" of the ligand from the complex, to prevent the ligand from escaping the binding site.[20] Similarly, in ref 22, a flat-bottomed harmonic well potential was employed to reduce the mobility of the ligand. Any such restraints need to be accounted for in terms of its free energy contribution, as pointed out by several workers.[26−28]

Gilson et al. carried out the most detailed analysis of the statistical−thermodynamical basis of the DAM.[27] The solution suggested by Gilson et al. (as well as by Roux and co-workers[26] and Wang and Hermans[28]) consists of breaking up eq 2b as follows

$$(PL)_{H_2O} \xrightarrow{\Delta A_{II}^{*}} (P)_{H_2O}\cdots(L)_{gas} \xrightarrow{\Delta A_{r}} (P)_{H_2O} + (L)_{gas} \tag{4}$$

The intermediate $(P)_{H_2O}\cdots(L)_{gas}$ is a hypothetical state where the interactions between L and P (as well as all ligand−solvent interactions) have been turned off (i.e., L is decoupled from the system in terms of physical interactions), but a restraint that holds L in a position and orientation resembling the native bound state has been introduced; the fact that the restraint is chosen to correspond to the bound position usually simplifies the computational problem, though alternative restraints could be used. The calculation of ΔA_{II}^{*} is thus free from the sampling problems occurring in DAM. To make clear the difference with respect to the original DAM of Jorgensen et al.,[16] Gilson et al. refer to this approach as *the double-decoupling method* (DDM).[27] In addition to solving the sampling problem at the "free" end point, the DDM leads to standard binding free energies with a clearly defined reference state; this result is described in the Theory section.

Of course, the original problem has been replaced in the DDM by that of calculating ΔA_{r}, and this work focuses on its correct and reliable determination. The calculation of this free energy difference by computer simulations is nontrivial, since, in the

absence of other interactions between the protein and the ligand, the restraint energy terms determine the position and orientation of the ligand relative to the protein. Removing one or more of the restrained degrees of freedom leads to the difficulties analyzed in the context of bonded energy terms by Boresch and Karplus.[30,31] Straightforward attempts to compute the free energy cost of adding/removing bond stretching and angle bending interaction terms in a simulation lead to divergent results, since it is necessary to have a criterion for deciding at which point a bond (or a bond angle) is broken. While one can circumvent this problem,[30] it is best to compute such free energy contributions analytically, as has been shown in an earlier analysis.[30,31] Thus, to make the DDM approach (i.e., eq 4) useful, in practice, the restraint needs to be chosen in such a manner that its free energy cost (ΔA_r) can be computed analytically. Such a restraint (or really, a set of restraints) should fulfill several criteria: (i) The analytical expression for ΔA_r should be theoretically sound and, if not exact, of sufficient accuracy for the problem under consideration. (ii) Both the position and orientation of the ligand need to be restrained *relative* to the receptor. (iii) While the free energy cost of the restraint (ΔA_r) needs to be computed analytically, the restraint (ΔA_{II}^*, cf. eq 4) is introduced by means of a molecular dynamics (MD) simulation. Thus, the restraint energy terms should be suitable for MD calculations.

Several applications of the DDM have been published.[26,28,32−34] However, the restraints used have not satisfied all of the above requirements. For example, Gilson et al.'s derivation corresponds to the use of a square-well potential for the restraint.[27] Accurate evaluation of ΔA_{II}^* for a square-well type potential by means of MD simulations is not straightforward, though a flat-bottom double harmonic well could be used.[22,35] Both square-well and flat-bottom harmonic well potentials limit the decoupled ligand to move beyond a spherical or ellipsoidal region. Depending on the system studied, this may or may not be enough to avoid the sampling problems of the original DAM. Similarly, several workers have restricted the position of the ligand, but not its orientation,[26,32,33] which is expected to be insufficient for obtaining accurate results in many cases. The original application of the versatile *body restraint algorithm* (BRA)[28] used absolute restraints rather than restraints relative to the receptor. Finally, to the best of our knowledge, only a single study published to date used such relative restraints.[34] An alternative to the DDM, which apparently avoids the need to explicitly compute ΔA_r, was suggested by Helms and Wade.[36]

The above brief review of the DDM and its existing applications makes clear that a closer look into methodological and practical aspects of DAM/DDM calculations is warranted. This is particularly true because ligand binding is a very important problem and the speed of present-day computers is such that statistically converged simulations could now be carried out, if a precise method for doing them were available. Our focus is on providing such a method, with particular attention to the second step of the DAM/DDM (eq 2b), since, as just outlined, it is the computation of ΔA_{II} where the problems of the original DAM arise. Also, the standard state dependence of ΔA_{bind}^0 can be accounted for in this step, cf. refs 27 and 28. The core of the present study is a versatile set of restraints for use in DDM free energy calculations based on MD simulations. An important point is that the restraints are such that analytic expressions for their free energy contribution, ΔA_r, can be derived. The analysis also makes clear the role of the standard state in binding free energy calculations. We then verify numerically the correctness of the formulation and test the utility of the proposed set of restraints. In particular, we demonstrate explicitly that the overall binding free energy does not depend on the details of the restraint(s) used in the DDM calculation.

The remainder of the manuscript is organized as follows. The Theory section (section 2) derives an analytic expression for ΔA_r for a particular set of restraints and analyzes the role of the standard state in binding free energies. In sections 3 and 4, respectively, we present the technical details and the results of two sets of calculations. In the first, which treats the complex formed by benzene with the L99A mutant of T4 lysozyme,[37] we compare two methods of calculating ΔA_r for our set of restraints, thus verifying numerically the correctness of the analytic derivations. In the second, we present results of DDM binding free energy calculations for a simplified model of a real ligand−protein complex, which consists of tyrosine complexed to tyrosyl-tRNA-synthetase.[38,39] We conclude with a Discussion section (section 5) focusing on two topics. First, we compare our set of restraints and corresponding analytic correction ΔA_r to the analysis of Gilson et al.,[27] as well as to earlier approaches,[26] and in particular to the BRA method of Hermans and Wang.[28,34] We also analyze the calculations by Helms and Wade.[36] Second, we briefly examine the relation between ΔA_r and the contribution of the loss of translational and rotational entropy to ΔA_{bind}.[27,40−51] Finally, an Appendix provides an analysis of the BRA method; we include this because we were originally confused about its applicability and it seems useful to spare others from this difficulty.

2. Theory

The DDM to compute absolute binding free energies was outlined in the Introduction. Its key element is a hypothetical intermediate state in which the interactions between receptor and ligand have been turned off alchemically but where restraints keep the ligand in a position and orientation that resembles the native bound state. To obtain ΔA_{bind}^0, the free energy ΔA_r of this restraint relative to the fully decoupled state needs to be calculated. In section 2.1 we derive an analytical expression for ΔA_r of a versatile restraint designed for use in DDM calculations. Section 2.2 outlines the role of the standard state in DDM calculations.

2.1. Virtual Bond Algorithm. *2.1.1. Classical Statistical−Mechanical Preliminaries.* In classical statistical mechanics, using Cartesian coordinates $\mathbf{r} = \{x_1...x_{3N}\}$, the kinetic energy contribution to the partition function Q can be computed analytically. The integration over the conjugate momenta $\mathbf{p}$ for each of the N atoms in the system (consisting of solute and solvent) can be carried out immediately, and one obtains the familiar expression[29,52]

$$Q = Z\prod_{i=1}^{N}\Lambda_i^{-3} \tag{5}$$

where Z is the configurational partition function and $\Lambda_i = h/(2\pi m_i kT)^{1/2}$ is the de Broglie wavelength (k is Boltzmann's constant, h is Planck's constant, and m_i is the mass of atom i). The evaluation of the configurational integral (i.e., the configurational partition function), on the other hand,

$$Z = \int_V \mathrm{d}\mathbf{r}\, \exp(-U(\mathbf{r})/k_B T) \tag{6}$$

is, in general, possible neither analytically nor numerically for a complex system.[53,54] It is, however, possible to rewrite the configurational partition function of a single N-atomic molecule in solution in the form[41,55,56]

$$Z = \int dx_1 ... dx_{3N} \exp(-\beta W(x_1 ... x_{3N})) =$$
$$\int dX\, dY\, dZ\, d\xi_1\, d\xi_2\, d\xi_3\, dx'_1 ... dx'_{3N-6}\, |\mathbf{J}| \times$$
$$\exp(-\beta W(x'_1 ... x'_{3N-6})) = 8\pi^2 V \int dx'_1 ... dx'_{3N-6}\, |\mathbf{J}| \times$$
$$\exp(-\beta W(x'_1 ... x'_{3N-6})) = 8\pi^2 V \tilde{Z} \quad (7)$$

Here W is the potential of mean force (PMF) resulting from the (formal) integration over the solvent degrees of freedom (DOFs). The $\{x_1 ... x_{3N}\}$ denote the Cartesian coordinates of the N solute atoms. Since W does not depend on the six external DOFs, one can always carry out a coordinate transformation separating external $\{X, Y, Z, \xi_1, \xi_2, \xi_3\}$ and internal DOFs $\{x'_1 ... x'_{3N-6}\}$, as shown in the second line of eq 7. The value of the Jacobian determinant $|\mathbf{J}|$ depends on the details of the coordinate transformation. The external coordinates $\{X, Y, Z\}$ are the origin of the new coordinate system; they could be, for example, the coordinates of an arbitrarily picked atom, the center of mass (COM) of the solute, or the COM of a subset of atoms. The remaining three external coordinates $\{\xi_1, \xi_2, \xi_3\}$ describe the orientation of the new coordinate system; they could be, for example, the Euler angles or a unit vector describing an axis plus an angle of rotation about this axis. For any nonlinear molecule, integration over the external DOFs always gives $8\pi^2 V$ (third line of eq 7), where V is the volume of the system. The symbol $\tilde{Z}$ denotes the configurational integral in the $3N - 6$ internal DOFs. This separability of external and internal DOFs provides the basis for our virtual bond algorithm (VBA), as well as the BRA method developed by Hermans and co-workers (which is analyzed and discussed in section 5.2, as well as in the Appendix).[28,34]

2.1.2. Derivation of the VBA. To derive the VBA, we consider the protein−ligand complex in the context of the DDM approach and focus on the evaluation of ΔA_r. Figure 2 depicts the end point of step ΔA^*_{II} of a DDM calculation, which we denote as the CL ("cross-linked") state; that is, we are concerned with the second part of eq 4

$$(P)_{H_2O} \cdots (L)_{gas} \xrightarrow{\Delta A_r} (P)_{H_2O} + (L)_{gas} \quad (8)$$

The ligand L has interactions neither with the receptor P (native) nor with solvent, but it is held by a set of restraints (which are specified shortly) in a position and orientation resembling the native bound state. The potential energy function of the entire system consists of three terms

$$U = U_P(\mathbf{r}_P) + U_L(\mathbf{r}_L) + U_r(\mathbf{r}_P, \mathbf{r}_L) \quad (9)$$

the protein term U_P (which should be understood as a PMF that includes the interactions of the protein with the solvent), the ligand term U_L, and the restraints U_r. To complete the DDM thermodynamic cycle, eq 4 requires knowledge of ΔA_r, that is,

$$\Delta A_r = -kT \ln \frac{Z_P Z_L}{Z_{CL}} \quad (10)$$

where Z_P, Z_L, and Z_{CL} are the partition functions for the protein, the ligand, and the hypothetical cross-linked state. An analytical calculation of ΔA_r is possible when the restraint terms are chosen so that a factorization of the partition function,

$$Z_{CL} = Z_P \times Z_r \times Z_L \quad (11)$$

into contributions from the protein Z_P, from the ligand Z_L, and from the restraints Z_r is possible, although the restraints $U_r(\mathbf{r}_P, \mathbf{r}_L)$ connect the protein and the ligand DOFs (cf. eq 9).

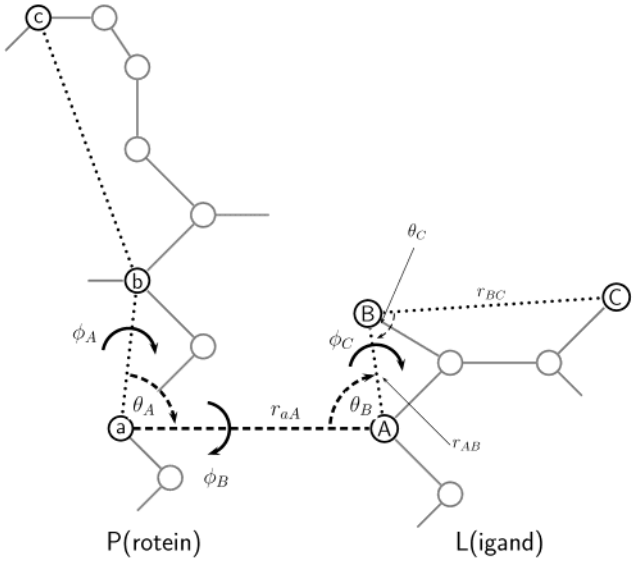

Figure 2. Definition of a cross-link restraining position and orientation of a ligand L relative to a receptor P. There are no other interactions between L and P. The cross-link consists of six (harmonic) restraint terms that act on three anchor atoms in the protein (a, b, c) and three anchor atoms in the ligand (A, B, C), respectively. As described in the main text, there is one restraint on the distance, r_{aA}, two bond angle restraints (θ_A, θ_B), and three dihedral restraints (ϕ_A, ϕ_B, and ϕ_C).

The Cartesian coordinate space of a system as depicted in Figure 2 consists of protein and ligand coordinates $\{\mathbf{r}_P, \mathbf{r}_L\}$; the solvent does not enter explicitly. We pick three (arbitrary) atoms A, B, and C in the ligand; the other $N_L - 3$ ligand atoms $\{\mathbf{r}_L'\}$ are not subject to any coordinate transformation. Two sets of coordinate transformations are applied to $\{\mathbf{r}_A, \mathbf{r}_B, \mathbf{r}_C\}$, the Cartesian coordinates of atoms A, B, and C. First, they are replaced by the relative coordinates $\mathbf{r}_{aA}$, $\mathbf{r}_{AB}$, $\mathbf{r}_{BC}$, where a is an arbitrary atom in the protein (see Figure 2). The Jacobian for this transformation is unity. The second set of coordinate transformations consists of introducing polar coordinates for each of the three relative coordinates; that is, $\mathbf{r}_{aA} = (r_{aA}, \theta_A, \phi_A)$, $\mathbf{r}_{AB} = (r_{AB}, \theta_B, \phi_B)$, and $\mathbf{r}_{BC} = (r_{BC}, \theta_C, \phi_C)$. For each transformation to polar coordinates a Jacobian determinant of $r^2 \sin\theta$ arises in the configurational integral. Six of the nine new polar coordinates (r_{aA}, θ_A, θ_B, ϕ_A, ϕ_B, and ϕ_C) are defined using the positions of three arbitrary protein atoms, labeled a (already used to define $\mathbf{r}_{aA}$), b and c; this is illustrated in Figure 2. (The other three coordinates (r_{AB}, r_{BC}, θ_C) only involve DOFs of the ligand; thus, they are part of Z_L and of no further concern.)

In the VBA, these six coordinates, which are the external DOFs of the ligand, are restrained by harmonic potentials $U_r(\zeta) = K_\zeta/2(\zeta - \zeta_0)^2$ to keep the ligand in a well-defined position and orientation relative to the protein, as mentioned earlier. Thus, Z_{CL} becomes

$$Z_{CL} = Z_P \int dr_{aA}\, d\theta_A\, d\theta_B\, d\phi_A\, d\phi_B\, d\phi_C\, r_{aA}^2 \sin\theta_A \sin\theta_B \times$$
$$\exp(-\beta(U_r(r_{aA}) + U_r(\theta_A) + U_r(\phi_A) + U_r(\theta_B) + U_r(\phi_B) + U_r(\phi_C)))$$
$$\int dr_{AB}\, dr_{BC}\, d\theta_C\, d\mathbf{r}'_L\, r_{AB}^2 r_{BC}^2 \sin\theta_C \times$$
$$\exp(-\beta U_L(r_{AB}, r_{BC}, \sin\theta_C, \mathbf{r}'_L)) \quad (12)$$

Equation 12 is the desired separation of Z_{CL}. The configurational integral of the protein Z_P can be separated out because the VBA coordinate transformations do not alter the protein

coordinates. The second integral in eq 12 is $\tilde{Z}_L$, that is, the configurational integral over the ligand internal DOFs after all external DOFs have been separated out; the symbol $\mathbf{r}'_L$ denotes the $N_L - 3$ coordinates of the ligand not subject to the coordinate transformations described above. The first integral in eq 12 over the six restrained external DOFs corresponds to Z_r; it can be evaluated using techniques developed by Herschbach, Johnston, and Rapp (HJR).[57] Assuming the validity of the rigid rotator (RR) approximation, the Jacobian factors $r_{aA}^2 \sin\theta_A \sin\theta_B$ can be replaced by their respective equilibrium values and taken out of the integral. The evaluation of Z_r then simplifies to six Gaussian integrals of the form $\int d\zeta \exp(-\beta K_\zeta/2(\zeta - \zeta_0)^2) = (2\pi kT/K_\zeta)^{1/2}$. Altogether, one obtains for the contribution of the restraints to Z_{CL}

$$Z_{CL}^{VBA} = r_{aA,0}^2 \sin\theta_{A,0} \sin\theta_{B,0} \frac{(2\pi kT)^3}{(K_r K_{\theta_A} K_{\theta_B} K_{\phi_A} K_{\phi_B} K_{\phi_C})^{1/2}} \quad (13)$$

Since the integral over these six external DOFs in the absence of the restraints gives $8\pi^2 V$ (cf. eq 7), one finds upon inserting eq 13 into eq 10 for ΔA_r with the VBA method

$$\Delta A_r^{VBA} = -kT \ln \frac{Z_P 8\pi^2 V \tilde{Z}_L}{Z_{CL}} = -kT \ln \left[\frac{8\pi^2 V (K_r K_{\theta_A} K_{\theta_B} K_{\phi_A} K_{\phi_B} K_{\phi_C})^{1/2}}{r_{aA,0}^2 \sin\theta_{A,0} \sin\theta_{B,0} (2\pi kT)^3} \right] \quad (14)$$

Interestingly, an equation corresponding to eq 14 was derived by DeVoe in 1969 (see eqs 37−40 of ref 55) for the purpose of estimating the free energy cost of cross-linking two proteins in the harmonic approximation, although no calculations were described by him.

The derivations leading to eq 14 have ignored symmetry considerations. If a molecule is symmetric, its configurational partition function contains an additional factor $1/\sigma$, where σ is the so-called symmetry number.[52,58] Thus, the presence of symmetry in, for example, the protein and/or the ligand leads to an additional contribution

$$\Delta A_{r,sym} = -kT \ln \frac{\sigma_{P\cdots L}}{\sigma_P \sigma_L} \quad (15)$$

Although in most cases no symmetry is present, there are exceptions;[28,45] for example, benzene, which is used in the model calculations reported in sections 3.1 and 4.1, has a symmetry number of 12.

Equation 14 for ΔA_r is the central result of this section. Its derivation relies on the separability of six external DOFs from the remaining $3N - 6$ internal DOFs, which is always possible (cf. eq 7). An essential part of the present method is that no further knowledge about Z_P and $\tilde{Z}_L$ is required; both cancel from eq 14. The particular simple form of eq 14 relies on (i) the validity of the RR approximation and (ii) the independence of the six external DOFs from intra- and intermolecular nonbonded interactions. The latter condition is automatically met since the cross-link restraint terms are the only terms affecting these six DOFs. Concerning the first condition, we can always choose sufficiently high force constants so that the RR approximation is fulfilled. In this case the equilibrium values can be substituted for the Jacobian factors r_{aA}, θ_A, and θ_B, and hence, these Jacobian factors can be taken out of the configurational integral. However, even if this were not possible, Z_r can still be computed. As seen from eq 13, Z_r factors into the product of six one-dimensional integrals, which can be computed either analytically (integrals in r_{aA}, ϕ_A, ϕ_B, and ϕ_C) or numerically (integrals in θ_A and θ_B).

2.1.3. Alternative Calculation of ΔA_r. The derivation of ΔA_r just presented makes clear that the free energy contributions of carefully chosen cross-links (such as the one in Figure 2) are independent of the free energy of the solvated protein ($-kT \ln Z_P$) and the decoupled ligand ($-kT \ln \tilde{Z}_L$). This makes it possible to compute ΔA_r by the rigid rotator harmonic oscillator (RRHO) approximation. In the RRHO approach the partition function Q of a polyatomic molecule is approximated by changing from Cartesian coordinates to a coordinate set consisting of the coordinates of the COM, the Eulerian angles specifying the orientation of the principal axes, and $3N - 6$ normal coordinates ν_i, as well as the conjugate momenta of these new coordinates.[52,57,58] The partition function is given by[52] $Q_{RRHO} = Q_{tr} \times Q_{rot} \times Q_{vib}$; that is, the partition function has been factored into translational (Q_{tr}), rotational (Q_{rot}), and vibrational (Q_{vib}) terms.[45,58] It follows that in the RRHO approximation the free energy of a polyatomic molecule is given by

$$A_{RRHO} = -kT \ln Q_{RRHO} = A_{tr} + A_{rot} + A_{vib} \quad (16)$$

The explicit expressions for the translational, rotational, and vibrational free energy contributions can be found in textbooks[52,58] or ref 45; the expressions are also given in the Supporting Information. Since ΔA_r is independent of the free energy of the solvated protein for the cross-links depicted in Figure 2 (as shown above), one can compute it according to

$$\Delta A_r = A_{RRHO}^{P} + A_{RRHO}^{L} - A_{RRHO}^{P\cdots L} \quad (17)$$

where all quantities are evaluated in the gas phase. The label RRHO indicates that the free energies of the protein, the ligand, and the cross-linked system are computed using eq 16. If symmetry needs to be considered, the correction is given by eq 15.[52,58] Although eq 17 is computationally much more involved (it requires *three* normal mode calculations—one for the protein P, one for ligand L, and one for the cross-linked complex P···L), it is nevertheless useful. First, it can be employed to verify numerically the correctness of eq 14; this is done in sections 3.1 and 4.1. Second, it can, in principle, be used for any combination of restraints. When doing so, however, one has to be careful concerning the number and type of DOFs that are subject to restraints (see the next subsection).

2.1.4. Additional Comments and Caveats. The particularly simple form of ΔA_r (eq 14) is a direct consequence of restraining exactly the six external DOFs of the ligand. One may be tempted to use additional restraints, for example, to add a second (pseudo) bond between a protein and a ligand atom to the six existing restraints. However, in this case one restrains more than just the six external DOFs and the simplicity of eq 14 is lost. One can still proceed by using eq 17. Unfortunately, complications resulting from restraining more than the six external DOFs reach deeper. To illustrate this complication, we consider as an artificial (extreme) example an intermediate state in a DDM calculation where the ligand is held by two sets instead of one set of restraints, each set being of the type shown in Figure 2. Assume that the two sets of restraints use different sets of anchor atoms in the protein, {a, b, c} and {d, e, f}, respectively. The simultaneous use of two cross-links results in the protein being influenced by the ligand attached to it.

Harmonic restraint potentials are functionally equivalent to the energy terms used to describe bonded interactions in many widely used force fields.[59−61] Thus, the presence of two sets of restraints implies that the restrained protein−ligand complex contains a ring. The ring resulting from the dual restraints leads to (at least indirect) interactions between the anchor atoms {a, b, c} and {d, e, f} that are not present in the native protein. Thus, this choice of the restraints violates the condition that the protein should not be influenced by the ligand in the artificial intermediate state. In other words, motions of the ligand couple with motions of the protein, and the behavior of the protein with a ligand connected to it in this manner is expected to be different from that of the protein without a ligand. Thus, when using eq 17 to compute ΔA_r in this case, one should keep in mind that ΔA_r obtained in this manner will not be exact. Related problems can arise in connection with the treatment of dummy atoms; cf. refs 62 and 63. Thus, when planning a DDM calculation, one should keep in mind that there are restrictions on the restraints that can (should) be used. Nevertheless, it may sometimes be useful to introduce additional restraints. An example might be a long, flexible ligand, such as a long alkyl chain or a peptide. Since eq 14 allows for only three anchor atoms in the ligand, one (or more) segment of such a ligand may remain extremely mobile, which is likely to slow the convergence of the calculation of ΔA^*_{II} considerably. The mobility of such a ligand could, however, be reduced by applying additional restraints on its dihedral angles.[28,35]

In the analysis so far we have stressed the need to compute ΔA_r analytically because the removal of the restraints depicted in Figure 2, which are functionally equivalent to bond stretching or angle bending energy terms, raises difficulties in standard free energy simulations.[31,64] In the central step of a DDM calculation, which is the evaluation of ΔA^*_{II}, these restraints are added to the system during a MD simulation. Consequently, one needs to examine why it is possible to compute ΔA^*_{II}, which effectively involves creating new bonds and bond angles, by computer simulations in a straightforward way. We consider the starting point and the end point of such a simulation. Initially, the ligand is attached to the protein by the native nonbonded interactions, which localize its position and orientation. As they are alchemically switched off, restraints which "take over" the task of restraining the position and orientation of the ligand are switched on. Thus, during all of step ΔA^*_{II} the position and orientation of the ligand relative to the protein remain determined. The free energy of such alchemical transmutations *can* be obtained from computer simulations. As pointed out earlier, there are several similarities between DAM/DDM calculations and the correct treatment of bonded interactions in free energy simulations. Step ΔA^*_{II} is analogous to the replacement of an angle bending term by a Urey−Bradley type energy term in the 1−3 distance. Since the angle bending degree of freedom remains restricted throughout the simulation, none of the sampling problems associated with an ideal-gas atom-like end state are encountered.[31] In step ΔA_{II} of the original DAM, on the other hand, the protein−ligand interactions are simply turned off, and the calculation is expected to fail to converge because of the associated sampling problems. The essential problem of step ΔA_{II} of the original DAM is completely analogous to the practical difficulties of turning off a bond length, angle bending, or Urey−Bradley type energy term.[31]

2.2. Standard State Dependence in Calculations of Binding Free Energies. *2.2.1. Thermodynamical and Statistical−Mechanical Considerations.* The chemical potential $\mu_{i,sol}$ of a molecule i in (aqueous) solution is given by

$$\mu_{i,sol} = \mu^0_{i,sol} + kT \ln \frac{\gamma_i c_i}{c^0} \tag{18}$$

where $\mu^0_{i,sol}$ is the standard chemical potential of i at the standard concentration c^0, c_i is the actual concentration, and γ_i is the activity coefficient. At the conditions we are concerned with (low concentration, atmospheric pressure), it is safe to assume $\gamma_i \approx 1$, which is done from now on,[27,65] although this is not necessarily appropriate for conditions in the cell.[66,67]

For a chemical reaction

$$\nu_A A + \nu_B B + ... \rightarrow \nu_R R + \nu_S S + ... \tag{19}$$

the equilibrium condition

$$\sum_i \nu_i \mu_{i,sol} = \nu_R \mu_{R,sol} + \nu_S \mu_{S,sol} + ... - \nu_A \mu_{A,sol} - \nu_B \mu_{B,sol} - ... = 0 \tag{20}$$

leads immediately to the usual expression for the standard free energy of the reaction, ΔA^0,

$$\Delta A^0 = \sum_i \nu_i \mu^0_{i,sol} = -kT \ln K^c \tag{21}$$

with

$$K^c = \frac{c_R^{\nu_R} c_S^{\nu_S} ...}{c_A^{\nu_A} c_B^{\nu_B} ...} (c^0)^{-\Delta\nu} \tag{22}$$

where $\Delta\nu = \nu_R + \nu_S + ... - \nu_A - \nu_B - ....$[29,68] The quantities of interest are ΔA^0 or K^c, and eq 22 demonstrates the well-known fact that (except in the special case $\Delta\nu = 0$) both depend on the chosen standard state; for example, $c^0 = 1$ mol/L and, in eq 18, c_i is measured in the same units (i.e., in mol/L). In particular, it follows that for the binding processes, eq 1, the fundamental quantity is the standard free energy change in the reaction, ΔA^0_{bind}, whose numerical value depends on the standard state.[29,52] This suggests that care in comparing calculated and experimental values is required to take account of the standard state used in the latter.

The same result can also be obtained by statistical−mechanical considerations. In the canonical ensemble $\mu_{i,sol}$ is given by[52]

$$\mu_{i,sol} = A_{i,N} - A_{0,N} = -kT \ln \frac{Q_{i,N}}{Q_{0,N}} \qquad \text{with } N \rightarrow \infty \tag{23}$$

The notation i, N and 0, N, respectively, in eq 23 means one or zero molecules i in N molecules of solvent (water); that is, $Q_{0,N}$ is the partition function of pure solvent. The symbols A and Q have their usual meaning of Helmholtz free energy and partition function (cf. eq 5). Comparing eq 23 with eq 18, one sees that $\mu^0_{i,sol}$ is given by

$$\mu^0_{i,sol} = -kT \ln \frac{c_i}{c^0} \frac{Q_{i,N}}{Q_{0,N}} = -kT \ln \frac{Q_{i,N}}{c^0 V Q_{0,N}} \tag{24}$$

where, in the last step, use has been made of the relation $c_i = 1/V$, since in molecular terms the concentration is given by the number of molecules, which is one, divided by the available volume V. Similarly, c^0 can be expressed as $c^0 = 1/V^0$, where V^0 is the standard volume (e.g., 1 l or 1 Å^3). One thus obtains

for the standard free energy of the reaction

$$\Delta A^0 = \sum_i \nu_i \mu^0_{i,\mathrm{sol}} = \frac{Q_{\mathrm{R},N}^{\nu_\mathrm{R}} Q_{\mathrm{S},N}^{\nu_\mathrm{S}} \cdots}{Q_{\mathrm{A},N}^{\nu_\mathrm{A}} Q_{\mathrm{B},N}^{\nu_\mathrm{B}} \cdots} (c^0 V Q_{0,N})^{-\Delta\nu} = \frac{Q_{\mathrm{R},N}^{\nu_\mathrm{R}} Q_{\mathrm{S},N}^{\nu_\mathrm{S}} \cdots}{Q_{\mathrm{A},N}^{\nu_\mathrm{A}} Q_{\mathrm{B},N}^{\nu_\mathrm{B}} \cdots} \left(\frac{V}{V^0} Q_{0,N}\right)^{-\Delta\nu} \quad (25)$$

depending on whether one prefers to use a standard concentration c^0 or a standard volume V^0. Equation 25 is the statistical mechanical counterpart of eq 21, with $Q_{\mathrm{A},N}$ defined as in eq 22. It is clear that eqs 21 and 25 have the same standard state dependence.

2.2.2. Application to the Calculation of Binding Free Energies. Specializing to the case of interest, binding of a ligand to a receptor (eq 1), eq 25 simplifies to

$$\Delta A^0_{\mathrm{bind}} = -kT \ln \frac{V Q_{\mathrm{PL},N} Q_{0,N}}{V^0 Q_{\mathrm{P},N} Q_{\mathrm{L},N}} \quad (26)$$

To make clear the dependence on a standard state V^0 (c^0), we write $\Delta A^0_{\mathrm{bind}}$ instead of just ΔA_{bind}. As already mentioned at the beginning of section 2.1, in classical statistical mechanics one can always separate out the kinetic energy contribution to Q, which cancels in the overall reaction, since the number of particles is conserved. Employing eq 5, denoting the number of atoms in the protein by N_P and that in the ligand by N_L, and noting that there are $3N_\mathrm{W}$ water atoms, one finds

$$\Delta A^0_{\mathrm{bind}} = -kT \ln \frac{V \prod_{i=1}^{N_\mathrm{P}+N_\mathrm{L}+3N_\mathrm{W}} \Lambda_i^{-3} Z_{\mathrm{PL},N} \prod_{i=1}^{3N_\mathrm{W}} \Lambda_i^{-3} Z_{0,N}}{V^0 \prod_{i=1}^{N_\mathrm{P}+3N_\mathrm{W}} \Lambda_i^{-3} Z_{\mathrm{P},N} \prod_{i=1}^{N_\mathrm{L}+3N_\mathrm{W}} \Lambda_i^{-3} Z_{\mathrm{L},N}} = -kT \ln \frac{V Z_{\mathrm{PL},N} Z_{0,N}}{V^0 Z_{\mathrm{P},N} Z_{\mathrm{L},N}} \quad (27)$$

Since the atom masses of the species involved in the numerator and the denominator cannot change, all de Broglie wavelengths, Λ_i, cancel, as indicated in the expression on the right. Finally, applying eq 7 to Z_P, Z_L, and Z_PL, one can write eq 27 also as

$$\Delta A^0_{\mathrm{bind}} = -kT \ln \frac{\tilde{Z}_{\mathrm{PL},N} Z_{0,N}}{8\pi^2 V^0 \tilde{Z}_{\mathrm{P},N} \tilde{Z}_{\mathrm{L},N}} \quad (28)$$

Except for two small differences, eq 28 is eq 13 of Gilson et al.[27] First, we use the canonical ensemble and obtain ΔA^0 while they use the *NPT* ensemble and calculate ΔG^0, which has the additional term $P^0 \Delta \bar{V}_{\mathrm{PL}}$. Second, we have ignored symmetry contributions; that is, we have omitted the factor $\sigma_\mathrm{P}\sigma_\mathrm{L}/\sigma_{\mathrm{PL}}$, cf. eq 15.[45,52,58]

The standard state dependence of $\Delta A^0_{\mathrm{bind}}$, although a textbook result,[29,52] is often overlooked; see the discussion in ref 69. One sees immediately that the original DAM (eq 2) neglects this standard state dependence; that is, neither V^0 nor c^0 appear in eq 2. In principle, the standard state correction could also be applied to the DAM. It is proportional to $\ln V/V^0$, where V is the volume of the simulation system (see also below). However, adding this correction is meaningful only if the decoupled ligand has sampled all of V, which is not feasible in practice (cf. the Introduction).

Analyzing the standard state dependence (if any) of the three steps that make up the DDM (eqs 2 and 4), one finds

$$\Delta A^0_{\mathrm{I}} = -kT \ln \frac{Z_{\mathrm{L},0} Z_{0,N}}{Z_{\mathrm{L},N}} \quad (29)$$

$$\Delta A^{*,0}_{\mathrm{II}} = -kT \ln \frac{Z_{\mathrm{P}\cdots\mathrm{L},N}}{Z_{\mathrm{PL},N}} \quad (30)$$

$$\Delta A^0_{\mathrm{r}} = -kT \ln \frac{V^0 Z_{\mathrm{P},N} Z_{\mathrm{L},0}}{V Z_{\mathrm{P}\cdots\mathrm{L},N}} \quad (31)$$

where $Z_{\mathrm{L},0}$ is the configurational partition function of the ligand molecule in the gas phase (i.e., when the ligand is decoupled from the system). Combining eqs 29−31 as required by the DDM (i.e., $\Delta A^0_{\mathrm{bind}} = \Delta A^0_{\mathrm{I}} - \Delta A^{*,0}_{\mathrm{II}} - \Delta A^0_{\mathrm{r}}$) leads to $\Delta A^0_{\mathrm{bind}}$, as given by eq 27, thus validating the thermodynamic cycle underlying the DDM. One sees that ΔA^0_{I} and $\Delta A^{*,0}_{\mathrm{II}}$ do not have a standard state (V^0) dependence and that all of the standard state dependence has been shifted to the calculation of ΔA^0_{r}.

In section 2.1.2 an analytical expression for ΔA_r was derived. Using eqs 14 and 31, we find the corresponding standard state dependent correction,

$$\Delta A_{\mathrm{r}}^{\mathrm{VBA},0} = -kT \ln \left[\frac{8\pi^2 V^0}{r_{\mathrm{aA},0}{}^2 \sin\theta_{\mathrm{A},0} \sin\theta_{\mathrm{B},0}} \frac{(K_\mathrm{r} K_{\theta_\mathrm{A}} K_{\theta_\mathrm{B}} K_{\phi_\mathrm{A}} K_{\phi_\mathrm{B}} K_{\phi_\mathrm{C}})^{1/2}}{(2\pi kT)^3} \right] \quad (32)$$

which is identical to eq 14, except that the volume of the simulation system V has been replaced by V^0, the volume corresponding to the chosen standard state: for example, 1660 Å^3 on an atomic basis for a one molar standard state. Equations 26−30 demonstrate not only that the DDM approach avoids the sampling problems of the DAM but also that it provides a framework in which the standard state dependence of the binding free energy can be incorporated in a straightforward manner.

It may seem tempting to avoid the explicit calculation of ΔA_r by introducing restraints during the computation of ΔA_I, analogous to those needed in step ΔA^*_{II}, since the two free energy contributions might then be expected to cancel. There are two problems with such an approach. First, adding restraints during step I of a DAM/DDM calculation (eq 2a) is likely to be nontrivial. The restraints depicted in Figure 2 are applied to external DOFs which are not subject to other interactions. Any attempts to calculate the free energy of adding such restraints will lead to the same problems thoroughly discussed in connection with the role of bonded energy terms in free energy difference calculations.[31,62−64] In special cases, for example, when using a flat-bottomed harmonic well potential,[35,36] it may be possible to compute the free energy of adding the restraint by a MD simulation. However, even in this case one still has not accounted for the standard state dependence of $\Delta A^0_{\mathrm{bind}}$. Since there is no real need to introduce a restraint in step I of a DDM calculation, it seems best to compute ΔA_r explicitly and include the standard state dependence as shown in eq 32.

3. Methods

Having derived eq 14 and shown how it provides a framework for double decoupling (DDM) calculations, we proceed to verify the practical applicability of the theoretical development

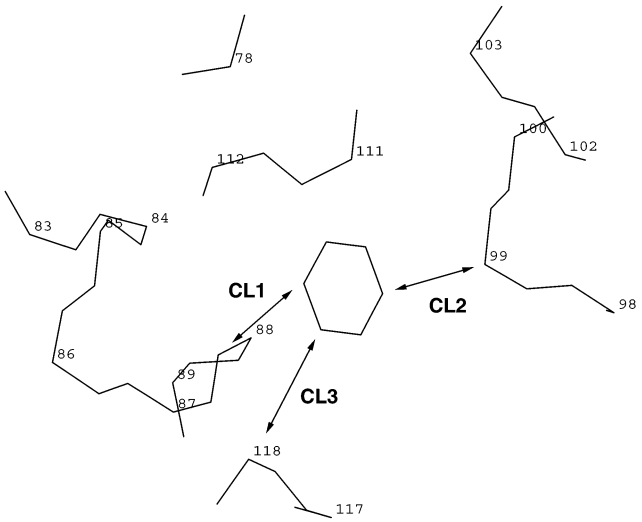

Figure 3. Backbone trace of the cavity created by the mutation Leu99 → Ala in T4 lysozyme (protein data bank code 181L). The small numbers denote residues by their C_α positions. The arrows indicate the three axes used to construct the three sets of restraints, CL1, CL2, and CL3, that are used to keep the ligand (benzene) in a well-defined position and orientation.

presented in the previous section. In this section we describe the methodology used to check numerically the correctness of eq 14 and then use it in a free energy simulation. The results of these calculations are presented in section 4.

3.1. Numerical Verification of Eq 14. The model system used in this first set of calculations is the complex of benzene with the L99A mutant of T4 lysozyme (protein data bank code 181L)[37] (see Figure 3), which was also used by Hermans and Wang when presenting their BRA method.[28] Rather than performing a complete calculation of the binding free energy, we just compute the free energy contribution ΔA_r (the second step of eq 4); that is, we calculate the difference in free energy between the free ligand in the gas phase (i.e., when it is completely decoupled) and the end point of the evaluation of ΔA_{II}^* (i.e., the state where all interactions between the ligand (benzene) and the receptor (T4 lysozyme) have been turned off alchemically, but the position and orientation of the ligand are restrained relative to the protein). ΔA_r was calculated for three different cross-links (referred to henceforth as CL1, CL2, and CL3) using both eqs 14 and 17; their position is shown in Figure 3 by arrows. Further, calculations in which the full protein was used are compared to calculations in which only a spherical shell surrounding the ligand is present. This second set of simulations is intended to confirm that the configurational partition function of the protein cancels in ΔA_r, so as to justify that it is correct to ignore solute−solvent interactions of the protein for the computation of ΔA_r. In fact, it suffices to use only a small part of the full protein, which considerably reduces the computational effort required for computing ΔA_r based on eq 17; in eq 14 there are no contributions from the protein.

The T4 lysozyme−benzene complex was set up using CHARMM version c28a2[70] and the *param19* polar hydrogen force field[71] with the EEF1 modified charges.[72] In all calculations a group based cutoff criterion was used; nonbonded interactions were smoothly switched off between 7 and 9 Å, and a distance-dependent dielectric function $\epsilon(r) = 1/r$ was used. We have used this simplified model to speed up the calculations so as to be able to do the converged numerical tests necessary for the present purpose. We, of course, realize that the model is unlikely to be adequate for an accurate free energy calculation.

TABLE 1: Details of the Cross-links Used in the T4 Lysozyme−Benzene Calculations[a]

		FULL	SPHERE
CL1[b]	r_{aA}	3.48	3.45
	θ_A	89.73	88.68
	θ_B	124.14	123.05
	ϕ_A	−166.40	−167.85
	ϕ_B	−74.83	−76.48
	ϕ_C	171.71	173.29
CL2[c]	r_{aA}	3.80	3.85
	θ_A	132.70	131.43
	θ_B	102.34	102.81
	ϕ_A	96.41	97.35
	ϕ_B	30.63	36.58
	ϕ_C	98.07	95.13
CL3[d]	r_{aA}	5.65	5.93
	θ_A	121.97	122.81
	θ_B	91.75	88.14
	ϕ_A	89.52	85.07
	ϕ_B	12.32	7.24
	ϕ_C	117.68	115.62

[a] All distances are in angstroms; all angles are in degrees. The meaning of the six degrees of freedom is depicted in Figure 2; the location of the three cross-links is shown in Figure 3. In all calculations the same force constants of 20 kcal/(mol Å^2) and 20 kcal/(mol rad^2), respectively, were used. [b] Atom c, C^{Tyr88}; atom b, C_α^{Tyr88}; atom a, N^{Tyr88}; atom A, C_1^{Bz}; atom B, C_6^{Bz}; atom C, C_5^{Bz}. [c] Atom c, C^{Ala98}; atom b, N^{Ala99}; atom a, C_α^{Ala99}; atom A, C_4^{Bz}; atom B, C_3^{Bz}; atom C, C_2^{Bz}. [d] Atom c, C^{Ser117}; atom b, N^{Leu118}; atom a, C_α^{Leu118}; atom A, C_2^{Bz}; atom B, C_3^{Bz}; atom C, C_4^{Bz}.

First, the position of benzene was optimized with 100 steps of steepest descent (SD) minimization, followed by 100 steps of adopted basis set Newton Raphson (ABNR) minimization, while holding the protein coordinates fixed. Then the full system was minimized with 3500 steps of SD and 20 000 steps of ABNR minimization. From this minimized structure (labeled henceforth FULL), an approximately spherical shell of approximately 12 Å radius surrounding the ligand was determined. All residues having one or more atoms further away than 14 Å from the center of geometry of benzene were deleted. The resulting system, henceforth referred to as SPHERE, consisted of 459 atoms instead of the 1621 atoms of the FULL system. This system was then energy minimized by 3500 steps of SD and 20 000 steps of ABNR minimization. This additional minimization is necessary, since the deletion of well over 1000 atoms results in considerable changes of interactions in the system. To prevent too large geometric changes, in this and all further calculations involving the SPHERE system, atoms lying more than 10 Å away from the center of geometry of the benzene ligand were restrained harmonically with a force constant of 1 kcal/(mol Å^2) to the position in the FULL system.

The FULL and SPHERE systems formed the basis for the introduction of three different sets of restraints of the type described in section 2.1.2 and depicted in Figure 2. The force constants of the six restraints were 20 kcal/(mol Å^2) and 20 kcal/(mole rad^2), for distances and angles, respectively, but the anchor atoms and hence the geometric equilibrium values of the harmonic energy terms varied significantly. Details are listed in Table 1, which also contains all data necessary to compute ΔA_r with eq 14. The positions of the restraints are indicated in Figure 3. Once the restraints were added, all nonbonded interactions between the benzene and T4 lysozyme were turned off. For the computation of ΔA_r by eq 17, normal mode calculations were carried out for T4 lysozyme, benzene, and the cross-linked benzene−T4 lysozyme complex using the VIBRAN module of CHARMM.[73] Prior to each normal mode calculation, the respective system (in particular, T4 lysozyme

without benzene and the cross-linked benzene−T4 lysozyme complex) was optimized further by 2500 steps of SD and 20 000 steps of ABNR minimization (unless the algorithm exited because convergence had been reached). This results in a root-mean-square (rms) energy gradient of less than 1×10^{-5}; there were no negative eigenvalues. For T4 lysozyme and the cross-linked complex, the rms difference for the backbone heavy atoms was 0.18 Å compared with the native complex with intact protein−benzene interactions; this is a result of the relaxation of the system following the removal (or decoupling) of benzene.

3.2. Application of Eq 14 to a Free Energy Simulation. In addition to verifying the correctness of eq 14 numerically, it is necessary to test its utility (and thus also that of the set of restraints depicted in Figure 2) in the context of free energy simulations. To achieve this goal, one has to show that identical binding free energies are obtained independent of the details of the cross-link used to restrain the ligand relative to the receptor after the native interactions have been removed. The system we studied is derived from the tyrosine−tyrosyl-tRNA-synthetase complex, which has been studied in great detail experimentally (e.g., refs 38 and 39) and which has also been the subject of computational studies.[7] To avoid the computational effort required for the realistic system,[7] we reduced its computational complexity to the essentials required for the present illustration and test.

Tyrosyl-tRNA-synthetase is a dimer; since the binding site is not located at the dimer interface, we studied only a monomer (as was done in ref 7). Further, we decided to leave out solvent, that is, to study the binding process in the gas phase. This implies $\Delta A_{\mathrm{I}} = 0$. Instead of having to deal with the full thermodynamic cycle of a DDM calculation (eqs 2 and 4), this permitted us to concentrate on obtaining the free energy difference ΔA^*_{II} without the influence of solvent. Omitting solvent also avoids a potential difficulty in real applications. At the end point of ΔA^*_{II} (ligand decoupled from protein and solvent but held by restraints in the binding pocket), water molecules could enter the binding site and occupy positions overlapping with the decoupled ligand. When carrying out backward simulations (restoring the native protein−ligand interactions), one has to be careful that none of these water molecules is trapped inside the binding pocket.

The protein was set up in CHARMM[70] using the *param19* polar hydrogen force field.[71] Since tyrosyl-tRNA-synthetase monomer is a large protein, only a part of the protein was included: All residues in the X-ray structure (PDB code 4TS1) which had any atom lying outside a 14 Å sphere centered about the C_α atom of the ligand were deleted; this resulted in an approximately spherical system with a radius of 12 Å. In all later calculations, atoms lying outside a spherical shell of 11 Å were restrained harmonically with a force constant of 1.0 kcal/(mol Å^2). The resulting system (including the ligand) consisted of 49 amino acids (414 atoms); 107 atoms were subject to position restraints. What was done here is analogous to a stochastic boundary simulation[7,74] in the absence of solvent. However, Nosé-Hoover thermostats were used to control the temperature instead of a region in which the system was treated as Langevin dynamics; cf. below.

To avoid sampling problems from the internal flexibility of the tyrosine ligand, two of its dihedral angles were modified. First, the force constant of the dihedral angle about the axis $C_\alpha-C_\beta$ was increased 10-fold from 1.6 to 16 kcal/mol. Second, we added a dihedral energy term for the atoms $C_\alpha-C_\beta-C_\gamma-C_{\delta 1}$ with a multiplicity of 1 and a force constant of 10 kcal/mol. This second dihedral prevents rotation of the aromatic ring.

TABLE 2: Details of the Cross-links Used in the Tyrosine−Tyrosyl-tRNA-Synthetase Model Calculations[a]

	L1A[b]	L1B[b]	L2A[c]
r_{aA}	5.10	5.20	5.20
θ_{A}	67.50	67.20	95.00
θ_{B}	84.50	88.00	96.00
ϕ_{A}	−113.00	−113.50	2.00
ϕ_{B}	−82.00	−81.80	−30.00
ϕ_{C}	−169.00	−170.50	124.00

[a] All distances are in angstroms; all angles are in degrees. The meaning of the six degrees of freedom is depicted in Figure 2; the two main links, L1 and L2, are also shown in Figures 4 and 5. [b] Atom c, N^{Pro39}; atom b, C^{Asp38}; atom a, $C_\alpha^{\mathrm{Asp38}}$; atom A, C_α^{Tyr}; atom B, C_β^{Tyr}; atom C, C_γ^{Tyr}. [c] Atom c, N^{Pro38}; atom b, C^{Phe37}; atom a, $C_\alpha^{\mathrm{Phe37}}$; atom A, C_γ^{Tyr}; atom B, C_β^{Tyr}; atom C, C_α^{Tyr}.

In principle, the free energy of this modification has to be computed. However, this correction is a constant offset that affects all binding free energies $\Delta A^0_{\mathrm{bind}}$ that were computed identically. The focus of this work is on the reproducibility and consistency of DDM calculations and not on the actual value of the free energy difference for the model process studied. Given all the simplifications of the system, the results are not expected to be realistic; therefore, we decided to omit the correction. Nevertheless, the calculation retains enough of the characteristics of a realistic application so that it can serve as a critical test of the methodology.

In all calculations, the nonbonded interactions were treated as follows. Both Lennard-Jones and electrostatic interactions were switched off smoothly between 7 and 9 Å. Interactions were truncated on the basis of group distances. A constant dielectric of 5 was employed. The nonbonded lists were updated heuristically on the basis of a 10 Å buffer region. As for the simplifications of the system studied, this choice of nonbonded options was motivated primarily by its reduction of computational effort.

Two sets of seven MD simulations (one set 300 ps and the other 500 ps in length) were carried out to investigate the stability of the reduced system. The tyrosine binding position changes significantly compared to the X-ray structure but was stable in the new structure. The binding pocket of tyrosyl-tRNA-synthetase is a deep cleft, which is, nevertheless, open to solvent. In the X-ray structure, the ligand is found deep inside the cleft and water entering the cleft prevents the rather hydrophobic tyrosine from escaping. Since we have no solvent in our system, the ligand moves out of the pocket to a new (artificial) binding mode. Once this binding mode has been reached, however, it represents a stable state, as verified by three additional 1 ns MD simulations started from this new binding position. On the basis of the last three calculations, possible candidates for the cross-link were identified. Table 2 summarizes the geometric parameters of the three links, L1A, L1B, and L2A, that were used in the actual free energy simulations (see below).

While L1A and L1B differ only marginally, L2A uses a different set of anchor atoms, both in the protein and in the ligand. A cross-link is fully characterized by the combination of geometric equilibrium parameters (summarized in Table 2) with a set of force constants for each of its six restraints. Different sets of force constants were used in the free energy calculations, ranging from 5 kcal/(mol Å^2) [kcal/(mol rad^2)] to 50 kcal/(mol Å^2) [kcal/(mol rad^2)]. For the remainder of this work, we adopt the following notation. For example, L1A-5 denotes the geometry L1A shown in Table 2 with all six force constants set to 5 kcal/(mol Å^2) [kcal/(mol rad^2)]. Similarly, L2A-50 is the combination of geometry L2A with six force

TABLE 3: ΔA_r of a Restrained Benzene−T4 Lysozyme Complex[a]

	Bz	T4	CL1	CL2	CL3
A_{tr}	−8.30	−13.18	−13.18	−13.18	−13.18
A_{rot}	−6.65	−15.86	−15.86	−15.85	−15.86
A_{vib}	10.98	2594.16	2600.08	2600.06	2599.48
$A_{tr} + A_{rot} + A_{vib}$	−3.97	2565.12	2571.04	2571.01	2570.44
ΔA_r[b]			9.88	9.86	9.29
ΔA_r[c]			9.88	9.86	9.29

[a] Results for the FULL system. All free energies are in kilocalories per mole. [b] Equation 17. [c] Equation 14.

TABLE 4: ΔA_r of a Restrained Benzene−Tr Lysozyme Complex[a]

	Bz	T4	CL1	CL2	CL3
A_{tr}	−8.30	−12.07	−12.08	−12.08	−12.08
A_{rot}	−6.65	−13.91	−13.91	−13.91	−13.91
A_{vib}	10.98	727.01	732.99	732.87	732.27
$A_{tr} + A_{rot} + A_{vib}$	−3.97	701.03	707.00	706.88	706.28
ΔA_r[b]			9.94	9.82	9.22
ΔA_r[c]			9.96	9.83	9.24

[a] Results for the SPHERE system. All free energies are in kilocalories per mole. [b] Equation 17. [c] Equation 14.

constants of 50 kcal/(mol Å^2) [kcal/(mol rad^2)]. In one case, labeled L1A-F, we used a combination of force constants: The distance restraint had a force constant of 4 kcal/(mol Å^2), the two angle restraints had a force constant of 8 kcal/(mol rad^2), and the three dihedral angle restraints had a force constant of 10 kcal/(mol rad^2).

All free energy simulations were carried out with the PERT module of CHARMM used in thermodynamic integration mode. Implementational limitations of the present PERT code make it impossible to realize the first step as it is written in eq 4. Consequently, this step (ΔA^*_{II}) was broken up into three substeps: (A) The cross-link was added while maintaining the native protein−ligand interactions. The restraints are introduced as bonded terms; thus, there is a change of nonbonded interactions in this step because of the change in connectivity (i.e., of excluded neighbors). In step B, with the cross-link already active, the protein−ligand interactions are turned off. In PERT this requires that the intraligand nonbonded interactions also be turned off, so that an additional step (C) is needed to restore the native interactions of the tyrosine ligand. While the results of steps A and B depend on the details of the cross-link, ΔA_C has to be calculated only once. The sum $\Delta A_A + \Delta A_B + \Delta A_C$ gives ΔA^*_{II}. Since for our simplified system (cf. above) $\Delta A_I = 0$, we have $\Delta A^0_{bind} = -\Delta A_{II} = -(\Delta A^*_{II} + \Delta A_r)$ from eqs 3 and 4. Table 5, which also contains the results, gives an overview of all the free energy difference calculations and provides some additional technical details. Each free energy difference reported is the average of at least five forward and five backward calculations. The shorthand used in the footnotes to Table 5 to describe the protocols has the following meaning. The time step in all MD simulations was 1 fs; the protocol name, for example, 465, is the number of MD steps times 1000; thus, the 465 protocol uses a total of 465 000 steps of MD. The first number in footnote *c* of Table 5 describing the 465 protocol is the number of MD steps used for equilibration at the respective initial λ value: $\lambda = 0$ for forward runs, $\lambda = 1$ for backward runs. All free energy protocols include data collection at both end points. The 465 protocol, the simplest protocol used, employs equally spaced grid points over the whole λ range $0 \le \lambda \le 1$. Each sampling point λ_i covers an area of integration of $\Delta\lambda = 0.05$. The first and last interval, corresponding to $\lambda = 0$ and $\lambda = 1$, always cover only $\Delta\lambda/2$, for example, 0.025 for the 465 protocol. The pair of numbers *n*/*m* following $\Delta\lambda$ means that at each λ_i (with the exception of the first λ value) *m* steps of MD were used for equilibration and *n* steps were used for data collection. As an example of a more complicated protocol, consider the 1687 protocol (footnote *d* in Table 5). 87 000 steps of MD were used for equilibration at the first λ value. Two grid spacings, $\Delta\lambda = 0.0222$ for the range $0 \le \lambda \le 0.9$ and $\Delta\lambda = 0.0033$ for the range $0.9 \le \lambda \le 1$, were used. At each λ value in the range $0 \le \lambda \le 0.9$, 7000 MD steps were discarded as equilibration, followed by 18 000 steps of data collection. In the second range, 5000 steps of equilibration were accompanied by 15 000 steps of data collection.

The free energy calculations for steps A and C used the standard PERT linear dependence of the hybrid potential energy function $U(\lambda)$ on the coupling parameter; that is,

$$U(\lambda) = (1 - \lambda)U_i + \lambda U_f \tag{33}$$

where the subscripts i and f denote the initial and final states of the alchemical transmutation, respectively. In all calculations pertaining to ΔA_B, soft-core potentials[75,76] were used to circumvent van der Waals end point problems. Finally, we note that it was imperative to use two Nosé-Hoover thermostats,[77] one for the protein and one for the ligand, to achieve converged simulations in step B. The target temperature of both thermostats was 300 K, and both thermal inertia parameters were set to 20 kcal s^2.

4. Results

4.1. T4 Lysozyme Results. The free energy, ΔA_r, corresponding to three different restraints for holding a decoupled benzene in a position and orientation relative to T4 lysozyme that resembles the native complex[37] was calculated using both eqs 14 and 17. The results for the FULL and SPHERE systems, as described in the Methods section, are summarized in Tables 3 and 4, respectively. The two tables list first the translational (A_{tr}), rotational (A_{rot}), and vibrational (A_{vib}) free energies of benzene by itself, T4 lysozyme by itself, and the three cross-linked systems (CL1, CL2, and CL3) in the harmonic approximation obtained with the standard approach[45] based on center of mass coordinates, three Eulerian angles describing the orientation, and a normal-mode analysis (cf. section 2.1.3, eq 16). The fourth line gives the sum of these three contributions; the fifth line contains ΔA_r for the cross-links computed according to $\Delta A_r = A^{CL}_{RRHO} - A^{T4}_{RRHO} - A^{Bz}_{RRHO}$ (cf. eq 17). These values should be compared with the entries in the last line, ΔA_r for the cross-link computed according to eq 14; for the values of the various terms appearing in eq 14, see section 3.1 (Table 1). We note that our results do not include the free energy corrections arising from the 6-fold symmetry of benzene.[52] The symmetry number for benzene is 12; therefore, from eq 15 one finds $\Delta A_{r,sym} = +kT \ln 12 = 2.14$ kcal/mol. However, since this term is identical in the two approaches, its omission is irrelevant for the present analysis.

For the FULL system the agreement between the two approaches (fifth and sixth line in Table 3) is perfect, thus clearly demonstrating the utility of eq 14. When looking at the ΔA_r values for the SPHERE system (Table 4), one notes that the overall ΔA_r results differ slightly (±0.06 kcal/mol) from those in Table 3. Since in the SPHERE system the protein atoms outside the sphere were deleted, the minimization following the generation of the sphere (cf. section 3.1) leads to a slightly different geometry of the reduced protein−benzene complex compared with that of the FULL system. Therefore, the cross-

TABLE 5: Details and Results of Free Energy Simulations for the Simplified Tyrosine−Tyrosyl-tRNA-Synthetase System[a]

	L1A-5[b]	L1A−F[b]	L1A-10	L1A-20	L1A-50	L1B-10	L1B-50	L2A-10	L2A-50[c]
protocol[d]	465[e]/1687[f]	465[e]/1556[g]	465[e]/1556[g]	1683[h]/1556[g]	1683[h]/1556[g]	465[e]/1556[g]	1683[h]/1556[g]	465[e]/1556[g]	1683[h]/1556[g]
ΔA_A	0.13 ± 0.04	0.33 ± 0.04	0.56 ± 0.06	1.11 ± 0.05	2.16 ± 0.09	0.49 ± 0.03	2.13 ± 0.11	1.26 ± 0.06	32.85 ± 0.45
ΔA_B	61.22 ± 1.19	62.36 ± 0.97	62.68 ± 0.94	63.42 ± 1.09	64.02 ± 1.06	62.48 ± 0.92	63.61 ± 1.12	62.62 ± 1.62	32.70 ± 1.16
ΔA_C[i]	−20.40 ± 0.13	−20.40 ± 0.13	−20.40 ± 0.13	−20.40 ± 0.13	−20.40 ± 0.13	−20.40 ± 0.13	−20.40 ± 0.13	−20.40 ± 0.13	−20.40 ± 0.13
ΔA_r^0	−6.27	−7.10	−7.51	−8.75	−10.39	−7.48	−10.36	−7.44	−10.32
H_B[j]	1.46	0.63	0.52	1.10	1.25	0.84	0.67	0.69	0.92[k]
ΔA_{bind}^0[l]	−34.69 ± 1.20	−35.19 ± 0.98	−35.33 ± 0.95	−35.39 ± 1.10	−35.38 ± 1.07	−35.09 ± 0.93	−34.98 ± 1.13	−36.05 ± 1.63	−34.83 ± 1.25

[a] For additional details concerning the simulation protocols, see the main text. All free energy differences are in kilocalories per mole. All free energy differences obtained from computer simulations are accompanied by the sample standard deviation, $s = 1/(n-1)(\sum_1^n (x_n - \bar{x})^2)^{1/2}$. [b] A calculation is labeled according to link (L1 or L2) and possible subtype (A or B); the number after the hyphen denotes the force constants used for the cross-link restraints. For example, L1A-5 denotes the geometry L1A indicated in Table 2 with all six force constants set to 5 kcal/(mol Å^2) [kcal(mol rad^2)]. L1A-F denotes the use of different force constants for the individual terms; that is, the distance restraint had a force constant of 4 kcal/(mol Å^2), the two angle restraints had a force constant of 8 kcal/(mol rad^2), and the three dihedral angle restraints had a force constant of 10 kcal/(mol rad^2). [c] A different intermediate state between steps A and B with rescaled charges was used; see text (section 4.2). [d] Protocols for steps A and B, respectively; the first number indicates the number of time steps used to compute ΔA_A, and the second number indicates that for ΔA_B; see text (section 3.2). [e] 50 000 steps of initial equilibration. $0 \le \lambda \le 1$: $\Delta\lambda = 0.05$ with 5 000/15 000 steps. [f] 87 000 steps of initial equilibration. $0 \le \lambda \le 0.9$: $\Delta\lambda = 0.0222$ with 7 000/18 000 steps. $0.9 \le \lambda \le 1$: $\Delta\lambda = 0.0033$ with 5 000/15 000 steps. [g] 50 000 steps of initial equilibration. $0 \le \lambda \le 0.08$: $\Delta\lambda = 0.004$ with 6 000/12 000 steps. $0.08 \le \lambda \le 0.38$ $\Delta\lambda = 0.015$ with 6 000/12 000 steps. $0.38 \le \lambda \le 0.78$: $\Delta\lambda = 0.04$ with 6 000/12 000 steps. $0.78 \le \lambda \le 0.93$: $\Delta\lambda = 0.015$ with 6 000/12 000 steps. $0.93 \le \lambda \le 1$: $\Delta\lambda = 0.0035$ with 6 000/12 000 steps. [h] 110 000 steps of initial equilibration. $0 \le \lambda \le 0.955$: $\Delta\lambda = 0.0286$ with 10 000/15 000 steps. $0.955 \le \lambda \le 1$: $\Delta\lambda = 0.0033$ with 8 000/20 000 steps. [i] Result obtained from one set of calculations using the 465 protocol;[e] see text. [j] Hysteresis between the means of the forward and backward calculations in step B. [k] Combined hysteresis of steps A and B. [l] $\Delta A_{bind}^0 = -\Delta A_{II} = -(\Delta A_A + \Delta A_B + \Delta A_C + \Delta A_r^0)$, since $\Delta A_I = 0$.

link parameters, which are chosen on the basis of the optimized geometry of the protein−benzene complex, are slightly different in the FULL and the SPHERE systems (see Table 1). These small differences of the link geometries lead to the small differences in ΔA_r between the FULL and the SPHERE systems. In addition, for the SPHERE system (Table 4) there are small differences between the ΔA_r results computed with eq 17 (fifth line) and eq 14 (sixth line). We note that differences of about 0.02 kcal/mol are so small as to be irrelevant in practical applications. Aside from the small differences in ΔA_r resulting from the use of slightly different cross-links, the comparison of the FULL with the SPHERE results demonstrates that the protein contribution cancels from ΔA_r. The terms involving the protein are dramatically different for the FULL and the SPHERE systems (T4 and CL1−CL3 entries in Tables 3 and 4, respectively), since over 1000 atoms have been deleted in the SPHERE system compared with the FULL system. However, the differences are completely consistent for the protein T4 and the cross-linked systems (CL1−CL3); that is why they cancel in the overall result for ΔA_r.

Tables 3 and 4 demonstrate another important point that should not be overlooked when computing the free energy cost of a cross-link with the VBA method. The values of ΔA_r for CL1, CL2, and CL3 are not identical even though the same force constants are used in the three cases (cf. section 3.1). This reflects the influence of the Jacobian factors on ΔA_r, which is made explicit by eq 14. While the angles have relatively little effect, the length of the a−A bond has a significant effect on the results. This can be seen, for example, by comparing the CL1 with the CL3 results (r_{aA} = 3.48 Å versus 5.65 Å). A difference of approximately 0.6 kcal/mol ($\approx kT$), although still small, cannot be ignored.

4.2. Results of Free Energy Simulations. The results of the free energy simulations carried out for the simplified tyrosine−tyrosyl-tRNA-synthetase system (cf. section 3.2) are summarized in Table 5. As described in section 3.2, the goal of these calculations was to show that $\Delta A_{II}^* + \Delta A_r^0 = \Delta A_{II}$ (see eq 4) and, hence, that $\Delta A_{bind}^0 = -\Delta A_{II}$ are independent of the details of the cross-link. As described in more detail in section 3.2, three such cross-links, denoted L1A, L1B, and L2A were used (see Figures 4 and 5, as well as Table 2). A specific calculation

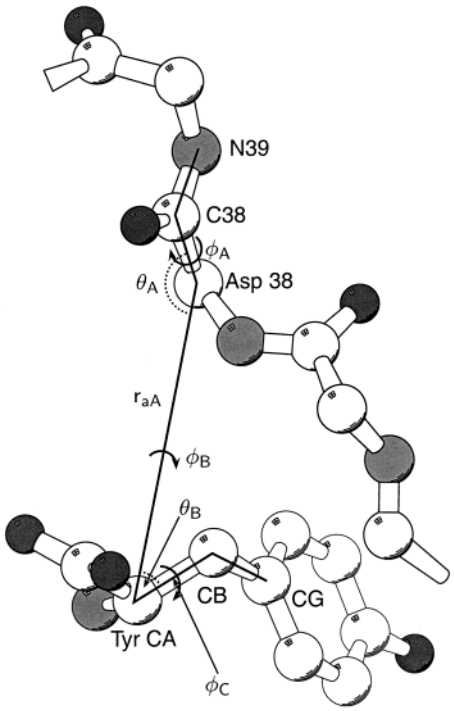

Figure 4. Detailed view of link family L1A/L1B which illustrates the atoms involved in the six degrees of freedom of the restraint. Carbon atoms are drawn in white, nitrogens in light gray, and oxygens in dark gray; no hydrogens are included in the drawing. Parts of the illustration were created using the program Molscript.[82] For further details concerning the link, see the main text and Table 2.

is referred to by the cross-link plus the value of the force constants; for example, L1A-10 refers to the calculations in which link L1A was used and the force constants were 10 kcal/(mol Å^2) and 10 kcal/(mol rad). Free energy simulations were used to obtain ΔA_{II}^*, while ΔA_r was evaluated analytically using eq 14. As outlined in section 3.2, the calculation of ΔA_{II}^* consists of three steps: A, addition of the cross-link restraint terms; B, removal of all nonbonded interactions involving the ligand; and C, correction for removing the intramolecular nonbonded interactions of the ligand in step B. The respective free energy differences are listed in the second, third, and fourth

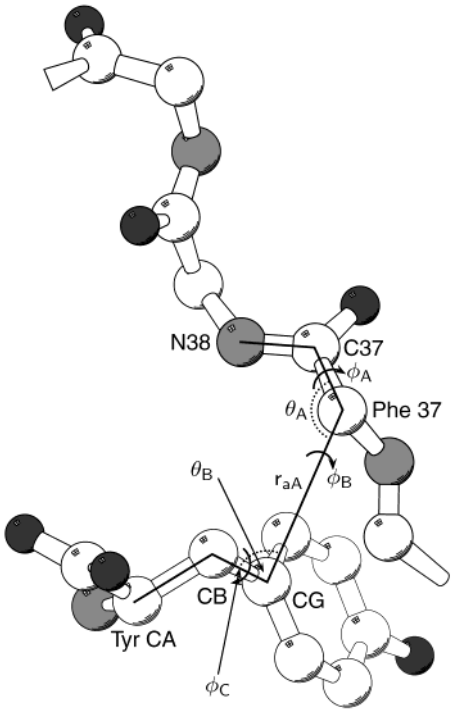

Figure 5. Detailed view of link L2A illustrating the atoms involved in the six degrees of freedom of the restraint. Carbon atoms are drawn in white, nitrogens in light gray, and oxygens in dark gray; no hydrogens are included in the drawing. Parts of the illustration were created using the program Molscript.[82] For further details concerning the link, see the main text and Table 2.

rows of Table 5. While ΔA_C is identical for all calculations, ΔA_A and ΔA_B depend on the details of the cross-link. The fifth line of the table gives the analytic correction for the free energy cost of the cross-link, ΔA_r^0, calculated for a standard state of 1 mol/L ($V^0 = 1660$ Å^3). The statistical uncertainty of the calculations is dominated by step B; the sole exception is set L2A-50, which is discussed in detail later. We also list the hysteresis H_B between the means of the forward and backward calculations of step B (sixth row in Table 5). This number is a measure of the reversibility of the calculations, which is an important test of the reliability of the binding constant calculation and the applicability of the chosen cross-links. In some of the original DAM applications, there were considerable problems with reversibility[17,18] because the ligand found a different binding mode in the backward calculation (i.e., it fell into a different, non-native local energy minimum).

The binding free energies ΔA_{bind}^0 (last line in Table 5) agree quite well; the average value of the nine results is −35.2 kcal/mol. This agreement should be compared with the variation of ΔA_r^0 as a function of (mainly) the different force constants; for example, for L1A, ΔA_r ranges from −6.27 kcal/mol (L1A-5) to −10.39 kcal/mol (L1A-50). Taking into account the statistical uncertainty (±1.7 kcal/mol), the results for $\Delta A_{bind}^0 = -\Delta A_{II}^*$ are shown to be independent of the details of the cross-link, as they are theoretically. The good agreement demonstrates the usefulness of the set of restraints depicted in Figure 2 for actual DDM calculations.

It is of some interest to analyze the results within the three families of restraints, L1A, L1B, and L2A. The largest number of calculations (five) was carried out for restraint L1A (see Figure 4 and Table 2). Comparing the standard deviation and, in particular, the hysteresis of step B (H_B) for the five sets of calculations, one sees that the best results (lowest standard deviation, lowest H_B) are obtained for the L1A-F and L1A-10 restraints. It seems that both too low force constants (L1A-5) and too high force constants (L1A-20, L1A-50) decrease the precision of the results. Too low force constants lead to a large conformation space accessible to the ligand. As a result, the sampling problems, which the restraints are intended to eliminate, are not completely removed. One might thus assume that higher force constants would lead to more precise results. However, the calculations with the larger 20 and 50 kcal/(mol Å^2) [kcal/(mol rad^2)] force constants have slightly higher standard deviations and, in particular, higher hysteresis than L1A-10 and L1A-F. Too high restraints can introduce a problem for the following reason. Although the conformation imposed by the restraint energy terms attempts to mimic the native binding mode, there will always be small differences between the native bound state and the cross-linked system. At the initial state of step B ($\lambda = 0$), both native protein−ligand interactions and cross-link restraints are present; as λ goes to 1, the native interactions are turned off. If strong restraints are applied, these terms will determine the conformation of the complex for most values of the coupling parameter. The native protein−ligand interactions will be dominant only for very small λ values. If there are any differences in binding mode for the native bound state and the cross-linked state held only by the restraints, then there will be a change in conformation of the ligand relative to the protein once the native interactions become too weak. Higher (too high) force constants may make this switch rather abrupt, and despite an overall rather elaborate and long free energy protocol, the sampling in this λ region may be insufficient, leading to the observed higher standard deviation and hysteresis.

Although restraint L1B is only a small variation of L1A (see Table 2), both results (L1B-10, L1B-50) have low hysteresis. The average geometry of the native binding mode is the result of a superposition of several conformations corresponding to local, energetic minima (data not shown). While the geometric parameters for L1A were chosen to match exactly one of these minima, L1B is the best fit to the average geometry. Apparently, this leads to the lower sensitivity of the quality of the result to the values of the force constants. The variation in behavior suggested the use of calculations with different anchor atoms in the protein and in the ligand, such as set L2A (see Figure 5 and Table 2). The binding free energy obtained with L2A-10 (−36.05 kcal/mol) has the largest deviation from the mean value over all calculations (−35.2 kcal/mol), though the difference is still small. The hysteresis is low, but this calculation also has by far the highest standard deviation. Several additional L2A-10 and L2A-50 calculations with even longer protocols failed to improve the precision of the results (results not shown). This prompted us to investigate the differences between the two sets of restraints (L1A/L1B versus L2A) in more detail. If one compares the restraints (see Figures 4 and 5), one sees that L1A/L1B reduces the mobility of the tyrosine backbone atoms but has little influence on the side chain (the aromatic ring). For L2A, the situation is reversed; this restraint lowers the mobility of the aromatic ring but leaves the backbone atoms essentially free to move. This conclusion was confirmed by studying the average geometry and fluctuations of the cross-linked system at the end point of step B (i.e., after all nonbonded interactions to the ligand had been turned off) (data not shown). L1A/L1B does have a comparatively high(er) mobility of the aromatic ring, while L2A has a high mobility of the backbone, including the charged NH_3^+ and CO_2^- groups. Inspection of the respective end points shows that for L1A/L1B the phenyl ring is located in a cavity; that is, despite its fluctuations there is little overlap with protein atoms. This is not the case for L2A; here the NH_3^+

and CO_2^- groups have a considerable overlap with protein atoms. While, in principle, this should not matter because of the use of soft-core potentials (cf. section 3.2), it seems that this is a problem and the convergence of the calculations (i.e., the evaluation of ΔA_{II}^*) is considerably slower with L2A compared to L1A/L1B.

To test whether the overlapping configurations involving the charged groups are responsible for the high standard deviation observed in the L2A-10 calculation, we carried out a set of calculations with a modified protocol (L2A-50). While the link itself was unchanged, we rescaled the ligand charges of the intermediate state between steps A and B. The end point of step A in the L2A-50 calculations was a cross-linked system in which the charges of the NH_3^+ and CO_2^- groups were modified so that the two groups were electrically neutral. (The rescaled charges were taken from the EEF1 force field.[72]) This explains the different value of ΔA_B and the much higher ΔA_A value in this calculation. The overall result of L2A-50 is certainly in better agreement with the average binding free energy than that of L2A-10. However, the standard deviation of L2A-50 remains relatively high. The poorer performance of L2A suggests that in practice some test simulations are required to select a satisfactory set of restraints.

An additional test that we tried was to use only a subset of the Figure 2 type restraints. We repeated the L1B-50 calculation (as described in Table 5) but, in one case, put only a single restraint on the distance between C_α of Asp38 and the C_α atom of the ligand and, in the second case, only held the three degrees of freedom r_{aA}, θ_A, and φ_A (cf. Figure 4). In both cases the forward calculations behaved normally. However, with only the distance restraint we did not obtain converged results in the backward simulations and a hysteresis of over 5 kcal/mol. By contrast, the binding affinity $\Delta A^0_{bind} = 35.98 \pm 1.27$ kcal/mol obtained with three restraints (r_{aA}, θ_A, φ_A) lies within the range of values found with all six restraints; also $H_B = 0.8$ kcal/mol is quite low (cf. Table 5). Holding these three degrees of freedom corresponds to restraining the position of the ligand but not its orientation. This shows that, at least for this system, position restraints are much more important than those on orientation.

5. Discussion

5.1. Significance of the Test Calculations. The results of section 4 demonstrate the usefulness of our approach to carry out DDM calculations. The data presented in section 4.1 prove the correctness of eq 14; the free energy simulations reported in section 4.2 are an example of how the analytic considerations leading to eq 14 can be fruitfully applied in the context of an actual application, even though the model is simplified by the absence of solvent. As one can see from Table 5, all computed binding free energies agree well within the statistical error bars, and the overall result ΔA^0_{bind} is independent of the details of the cross-link, as it must be. Choice and parametrization of the link can, however, influence the convergence of the results. Use of the second link, L2A, led to binding free energies with somewhat higher standard deviations. Further, two variations of the first link (L1A/L1B) were used. For L1A, too low and too high force constants led to results that exhibited a noticeable hysteresis between the forward and the backward calculations (see the H_B values, previous to the last line in Table 5). By contrast, both L1B calculations gave equivalent results. Although we use in some instances somewhat optimized thermodynamic integration protocols with a variable grid spacing (in particular, the 1556 protocol), no attempts were made to refine these protocols. Continuously adapted grid spacing protocols[62] may well improve convergence and, hence, lower standard deviations of the results. Further, very good results were obtained with the L1A-F protocol, in which force constants of different strength were used for the different restraint terms. Thus, it may be advantageous to choose the strength of the restraint terms on the basis of a detailed analysis of the flexibility of the ligand in the native bound state.

Although the VBA restraints provide considerable flexibility, problems may arise for long, highly flexible ligands. In principle, ΔA_r for more complicated (or additional) restraints can be calculated numerically using eq 17. Since it is not necessary to include the full protein in the normal mode calculation (see the FULL versus SPHERE results in section 4.1), system size is not a limiting factor. However, eq 17 ceases to be exact if more than the six external DOFs of the ligand are restrained (cf. the discussion of the use of two Figure 2 type restraints in section 2.1.4). Nevertheless, as pointed out there, situations may arise in practice where one has to weigh the benefits of tighter but redundant restraints that improve convergence of the MD based free energy simulations against the drawbacks of possibly inaccurate ΔA_r corrections because of redundancy. An alternative to the use of a more complicated and theoretically problematic set of restraints may be to make the ligand more rigid (in the limit of a rigid ligand, a Figure 2 type cross-link is always sufficient), as suggested by Hermans and co-workers.[28,35] Here problems may arise in correcting for the restraint(s) on the dihedral angle(s) of the ligand if the bound ligand adopts conformations that do not correspond to a conformational free energy minimum in the unbound state.

5.2. Comparison with Published Approaches. The calculation of free energies of binding (protein−ligand, protein−protein) is an important part of computational biology. Although a number of proposals for binding free energy calculations already exist in the literature, many of them appear to lack something in clarity or completeness; that is, the theoretical development is not spelled out and/or some aspects of the practical application of the methodology are not fully described. We have presented a method for determining the free energy of binding that is exact, in principle, and have demonstrated by test calculations that the method is useful for practical calculations. To make clear the complexity of the problem and to point out certain calculational difficulties, which can be avoided by using the present approach, we discuss some of the published proposals in light of the present development.

A landmark paper on ligand binding calculations was published a few years ago by Gilson and co-workers.[27] Their reformulation of the double annihilation method (DAM method) is expressed mathematically by eq 28 of their study.[27] The third and fourth terms on the right-hand side of eq 28 correspond in our notation to ΔA_r. Adapting the notation slightly to match the present work, these two terms can be written as

$$\Delta A_r^{\text{Gilson}} = -kT \ln \frac{8\pi^2 V^0}{V_I \xi_I} \tag{34}$$

where V_I and ξ_I are the integrals over the six degrees of freedom determining the position and orientation of the ligand in the cross-linked state. Equation 34 is based on the always possible separation of external and internal DOFs (cf. eq 7). The construction of the external ligand coordinates used by Gilson et al. is analogous to our approach. Our eq 14 can, therefore, be regarded as a specific and useful formulation for calculating V_I and ξ_I in eq 34, even though no specific "recipe" for doing

so is given in ref 27. The authors argue that eq 34 is appropriate for a square-well potential, which confines the ligand to a certain volume and orientation, but no details are presented to show how such a square-well potential is to be implemented in a molecular dynamics simulation.

Equation 34 was used, for example, in a study on the binding free energies of anthracyclines to DNA.[49] Reference 49 used continuum methods (Poisson−Boltzmann calculations plus solvent accessible surface computations) rather than MD based free energy difference simulations. Values for V_I and ξ_I were estimated either from a grid search based on energetic criteria or from a MD simulation in which the center of mass and the inertia ellipsoid of the ligand were monitored. Gilson and coworkers[78] also used the theoretical foundation presented in ref 27 to develop a method to compute the conformational free energy of ligands. A related technique to this so-called "mining minima" approach has been introduced by Kolossváry.[79] The utility of these methods[78,79] lies in the rapid *approximate* computation of relative binding free energies with implicit solvent models and the possibility of taking into account the conformational flexibility of the ligands.

In a detailed study analyzing the thermodynamic stability of water molecules in the bacteriorhodopsin proton channel, Roux et al.[26] used a harmonic potential with force constant K_b on the oxygen of the probe water molecule to restrict its mobility. No orientational restraints were applied. The correction for the restraint was found to be

$$\Delta A_r^{\text{Roux}} = -kT \ln \frac{V^0}{(2\pi kT/K_b)^{3/2}} \tag{35}$$

This expression has also been used in earlier work from Jan Hermans' group.[32,33] Equation 35 is obtained for a restraint potential of the form $U_r = (K_b/2)\Delta\mathbf{X}^2$, where $\Delta\mathbf{X}$ is the displacement vector from the reference position, which was a fixed point in space, since the protein (the bacteriorhodopsin channel) was not allowed to move; that is, there was no net translation or rotation of either the receptor or the complex. The restraint used in ref 26 has two limitations. First, as already mentioned, it only restrains the position but not the orientation of the ligand, which is likely to cause difficulties for larger ligands. Second, a general (set of) restraint(s) should be relative to a moving reference point.

Both points are addressed by the BRA method of Hermans and Wang.[28,34] Interestingly, the first application of the BRA used reference points and axes that were fixed in space,[28] similarly to earlier DDM calculations.[26,32,33] In a recent application of the BRA, Mann and Hermans[34] used reference points relative to the moving protein; however, since the ligand was a xenon atom, only the position restraint of the BRA was tested. Nevertheless, the BRA is a general framework to carry out DDM calculations, and it is of interest to provide a detailed comparison with our VBA approach. Since ref 28 appears not to prove the separability of the partition function of the cross-linked state, $Z_{CL} = Z_P \times Z_r^{BRA} \times \tilde{Z}_L$ (cf. eq 11), we provide a brief theoretical analysis of the coordinate transformations involved in the Appendix. The most striking difference is that the BRA requires (implicitly at least) the definition of the external DOFs of the protein, whereas the VBA leaves the protein DOFs completely untouched. Switching to the notation of eq 7, the BRA ties the external DOFs of the ligand $\{X_L, Y_L, Z_L, \xi_1^L, \xi_2^L, \xi_3^L\}$ to those of the protein $\{X_P, Y_P, Z_P, \xi_1^P, \xi_2^P, \xi_3^P\}$ by restraining the respective deviations $\Delta X_{PL} = X_L - X_P$ and so forth. This implies a choice of the reference points in the protein, so that, in some reference position (e.g., the X-ray structure of the protein−ligand complex), $\{X_P, Y_P, Z_P, \xi_1^P, \xi_2^P, \xi_3^P\}$ and $\{X_L, Y_L, Z_L, \xi_1^L, \xi_2^L, \xi_3^L\}$ coincide. In VBA one chooses six DOFs that define the position and orientation of the ligand *relative* to the protein. These are restrained to the values that they have in the reference position by applying, for example, internal coordinate restraints or by introducing (virtual) bonded terms; as outlined in section 2.1.2, we use harmonic potentials in the respective internal coordinate.

We have presented the VBA using atom positions as anchors in both the protein and the ligand. By contrast, in the BRA, linear combinations of actual (atomic) coordinates are employed; thus, the BRA may seem to be even more general than the VBA. However, the BRA *requires* such linear combinations of atomic coordinates for the reference/anchor points. In the BRA the reference DOFs $\mathbf{X}_{i,0}$, $\mathbf{e}_\theta$, and χ (defined by $\mathbf{e}_\chi$) have to be chosen in such a manner that the external DOFs of protein and ligand ($\{X_P, Y_P, Z_P, \xi_1^P, \xi_2^P, \xi_3^P\}$ and $\{X_L, Y_L, Z_L, \xi_1^L, \xi_2^L, \xi_3^L\}$) coincide (cf. above and the Appendix). Without linear combinations of atom positions, this match would be impossible in most cases. For example, Mann and Hermans report that $\mathbf{X}_{i,0}$ used to restrain xenon in binding position 1 is the center of mass of 13(!) protein atoms.[34] Xenon in binding position 2 requires a different reference position (which is not explicitly specified in ref 34). Although there is no protein atom at the site of the xenon atom, the nature of the position restraint used by the BRA requires a reference position that coincides with the position to which the xenon atom is restrained. By contrast, the VBA restraints can always be defined in terms of three protein atoms. In the T4 lysozyme xenon example (in principle), any three protein atoms could be used to define a reference position for the VBA position restraint. Thus, in practice, the apparent additional flexibility of the BRA (i.e., the use of linear combinations of atom positions instead of just atom positions) is in our opinion negated by the requirements for choosing the reference DOFs. In the context of the VBA, we cannot envision a situation where linear combinations of protein atom positions would be advantageous compared with atomic positions. Concerning the ligand, the VBA could easily be extended to use linear combinations of ligand atom positions as anchor points. All that is needed is an initial variable transformation from $\{\mathbf{r}_A, \mathbf{r}_B, \mathbf{r}_C, \mathbf{r}'_L\}$ to $\{\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C, \tilde{\mathbf{r}}'_L\}$, where $\{\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C\}$ are the linear combinations of ligand atoms to be used as anchor points, and $\tilde{\mathbf{r}}'_L$ denotes all other ligand coordinates. The Jacobian determinant arising from this transformation would be absorbed in $\tilde{Z}_L$. Then, one would proceed as in the standard VBA using $\mathbf{X}_A$, $\mathbf{X}_B$, and $\mathbf{X}_C$ instead of $\mathbf{r}_A$, $\mathbf{r}_B$, and $\mathbf{r}_C$. The practical implementation of such restraints, however, may be more complicated.

In a study exploring the hydration energy landscape of the active site cavity in cytochrome P450cam, Helms and Wade proposed an alternative to DDM calculations.[36] To determine the thermodynamically most favorable number of water molecules in the active site, several calculations were carried out in which the free energy of transferring (removing) one (additional) water molecule into (from) the active site was computed. This problem is completely analogous to the calculation of a binding free energy. Sampling problems were avoided by restraining the water(s) to be inserted into the cavity (corresponding to step II of the DAM/DDM cycle in eq 2) with a flat-bottomed harmonic well potential. Use of an analogous restraint in what we refer to as step I of the DAM/DDM cycle [i.e., when decoupling the ligand (in this case a water molecule) from solvent] avoids the need to compute ΔA_r explicitly (cf. section 2.2). Nevertheless, one still has to correct ΔA_{bind} for

the standard state of interest. In the calculations reported in ref 36, *m* water molecules were decoupled from interactions with the rest of the system while simultaneously turning on interactions with *n* water molecules initially in the gas phase; *n* and *m* differ by one. Helms and Wade find as the standard state correction (appendix of ref 36)

$$\Delta A_C^{HW} = -kT \ln(m/n) \tag{36}$$

Equation 36 is puzzling. First, it implies that the standard state correction depends on the number *m*, *n* of water molecules that are decoupled from and recoupled to the rest of the system. While carrying out the calculations in this manner may be advantageous from a methodological point of view, it does not change the fact that the overall process studied is the transfer of one water molecule into a binding pocket. For the latter, the standard state correction is given by $-kT \ln(V^0/V_{sim})$. It is clear that ΔA^0_{bind} of transferring one water molecule into the cavity of a protein depends on the number of water molecules already in the pocket. However, this is a consequence of the difference in the interactions and not of the standard state correction. Second, ΔA_C^{HW} is undefined for the net transfer of one water molecule into the cavity ($n = 1$, $m = 0$). Helms and Wade specifically exclude this case, but there is no physical justification why this process should be disallowed or should have to be treated differently.

5.3. ΔA_r versus Loss of Translational/Rotational Entropy upon Binding. A long-standing point of contention concerns the magnitude of the loss of translational and rotational entropy[27,40−51] minus the gain of vibrational entropy that occurs when a ligand binds to a macromolecule or two macromolecules form a complex. There is no experimental method for determining this quantity (e.g., the attempt to do this by Tamura and Privalov[80] actually measures a different quantity[81]), and the calculated values cover a wide range (from 15 to 100 cal/K), which appears difficult to explain solely on the basis of the differences among the systems (or processes) studied. A general analysis of this important question is beyond the scope of this work; we shall return to it in a subsequent paper (see also refs 27 and 48). However, we comment briefly on the relation between ΔA_r (eq 14 or eq 17) and the loss of translational and rotational entropy upon binding, ΔA_{loss}, sometimes also referred to as the cratic free energy.

During any binding or association process, one partner loses its translational and rotational degrees of freedom; in exchange, the complex (the newly formed molecule) has six additional vibrational degrees of freedom. It seems tempting to apply expressions for ΔA_r as derived in section 2.1 to compute (or at least estimate) the corresponding free energy difference ΔA_{loss}. Hermans and Wang, for example, essentially regard the ΔA_r values, which they compute with the BRA, as the cratic free energy contribution. However, one has to separate conceptually the computation of ΔA_r in DDM calculations and the unambiguous determination of the loss of translational and rotational entropy. Our results for the T4 lysozyme−benzene complex (section 4.1) show that ΔA_r depends on the details of the restraining potential. Analogous findings were obtained in the binding free energy calculations reported in section 4.2 (see the ΔA_r values in Table 5, which differ by more than 3 kcal/mol). In DDM calculations the restraining potential (e.g., a harmonic cross-link as proposed by us) is introduced to avoid convergence problems in the free energy simulations. Since the value of ΔA_r is a function of the choice of the parameters of the restraining potential, it will in general not be identical to ΔA_{loss}. To make this clear, we consider again the two steps involved in a DDM calculation (cf. also refs 27 and 28). First, the alchemical calculation is carried out; in this step the native interactions between ligand and protein are replaced by some restraining potential (whether this is done in one step or in three steps, as in this work, is not of importance). In the second step, the free energy cost of the restraint is computed. In the context of a DDM calculation, one aims to find a restraint which sufficiently reduces the mobility of the decoupled ligand to avoid the sampling problems that would result otherwise. However, the choice of the restraining potential influences the relative contribution of the two terms (as illustrated by the results of section 4.2). As the results show, one has considerable leeway concerning the choice of the restraint parameters, while the result of the overall computation remains independent of the restraint. For DDM simulations this is not a drawback, since insensitivity of a method to the choice of its parameters (the cross-link in our approach) is, in fact, a desirable property. However, the variability in ΔA_r makes clear that identifying it with ΔA_{loss} is not straightforward.

Referring to eq 34, the determination of ΔA_{loss} requires knowledge of the true values of V_I and ξ_I (i.e., the integrals over the six degrees of freedom determining the position and orientation of the ligand); see also Gilson et al. (left column on p 1063 of ref 27). Nevertheless, eq 14, which is a specialization of the general expression eq 34, may prove useful for estimating ΔA_{loss}. On the basis of a molecular dynamics simulation of the native bound state, one can, in principle, derive parameters for a Figure 2 type cross-link, which, when replacing the native protein−ligand interactions, leads to a ligand behavior (average position and orientation, as well as fluctuations about the mean values) that closely resembles the native bound state. Equation 14 evaluated with these parameters should lead to a meaningful value for ΔA_{loss}. To further test the adequacy of the results, one could fit Figure 2 type restraint parameters for several sets of anchor atoms in the protein and in the ligand. The variability in the values of ΔA_{loss} obtained in this fashion would serve as an error estimate.

6. Conclusion

We have presented an exact framework for computing binding affinities by molecular dynamics (MD) based free energy simulations. The core of our approach, which is an implementation of the double-decoupling method (DDM), is a versatile set of restraints that avoids the sampling problems present in the original method (the DAM). The proposed restraints (Figure 2) can be used in MD simulations to restrict both the position and orientation of the ligand when it no longer interacts physically with the receptor, and they are defined relative to the receptor rather than relative to a fixed point in space. The free energy cost ΔA_r for this set of restraints was derived analytically; the resulting expression (eq 14) was shown to be exact. Our approach accounts correctly for the standard state dependence of binding free energies. Both the correctness of eq 14 and the general applicability of the method were demonstrated by calculations for two model systems. The complex formed by benzene with the L99A mutant of T4 lysozyme was used to verify numerically the derived expression for ΔA_r. By calculations on a simplified tyrosine−tyrosyl-tRNA-synthetase complex, we showed that there are no problems in applying the method in actual MD simulations. In particular, we demonstrated that the calculated overall binding free energy is independent of the details of the auxiliary restraints.

In most earlier applications of the DDM, the restraints limiting the mobility of the ligand were not fully general. Often, only

the position was restricted,[26,32,33] and the first application of the fully general body restraint algorithm (BRA) used restraints with reference points fixed in space.[28] In a second application of the BRA, restraints relative to a moving protein were employed, but the ligand (a single xenon atom) was extremely simple. The VBA approach presented here provides an easy to use alternative to the BRA of Hermans and co-workers. Although the model systems used to numerically verify the methodology were deliberately kept as simple as possible from a methodological point of view, they represent the most general DDM calculation reported today (restraints on both the position and orientation of the ligand relative to the protein). Both the BRA and the approach developed here present versatile methods for avoiding sampling problems in step II of the DAM thermodynamic cycle (eq 2b) *and* for taking the standard state dependence of binding free energies into account. We have also provided a theoretical analysis of the BRA method (and proof of its correctness) that is not given in the original work.[28,34] Having available a framework for computing binding affinities that is, in principle, exact will hopefully lead to applications in biology and drug design for which a knowledge of the values of these quantities and the interactions involved is essential for our understanding.

Acknowledgment. We thank Jan Hermans for making clear to us that the BRA is in fact valid and for other helpful comments on the manuscript. M.K. thanks the National Institutes of Health for partial support of this research. S.B. ackknowledges helpful discussions with Othmar Steinhauser.

Appendix. Analysis of the BRA Method

While overall very similar to the VBA method, the BRA algorithm of Hermans and Wang (HW)[28] does use a different set of restraints. It is, therefore, necessary to prove whether these restraints can be factorized from Z_{CL}, as demanded by eq 11. In the following we adopt the notation of ref 28. The BRA uses three reference points $\mathbf{X}_{i,0}$, $\mathbf{X}_{j,0}$, and $\mathbf{X}_{k,0}$, and three anchor points in the ligand $\mathbf{X}_i$, $\mathbf{X}_j$, and $\mathbf{X}_k$. Each of these anchor points can be a ligand atom or a linear combination of a subset of ligand atoms, such as their COM. The application reported in ref 28 used absolute reference points (i.e., fixed points in space); however, in the most general case, the reference points can be protein atoms or linear combinations of a subset of protein atoms. For example, in a recent study by Mann and Hermans investigating the binding affinity of xenon atoms to the L99A mutant of T4 lysozyme,[34] the reference point $\mathbf{X}_{i,0}$, to which the position of the xenon atom was restrained, was the center of mass of several protein atoms. The following restraints are applied. Anchor point $\mathbf{X}_i$ is tied harmonically to the reference point $\mathbf{X}_{i,0}$. The axis $\mathbf{r} = \mathbf{X}_j - \mathbf{X}_i$ is restrained to be aligned with a unit vector $\mathbf{e}_\theta = |\mathbf{X}_{j,0} - \mathbf{X}_{i,0}|$ using the restraint function $U_{r,\theta} = K_\theta/2(1 - \cos\theta)$ with $\cos\theta = \mathbf{e}_\theta \cdot \mathbf{r}/r$. Finally, a harmonic restraint is applied to keep the angle χ between the two planes formed by the three reference points $\mathbf{X}_{i,0}$, $\mathbf{X}_{j,0}$, and $\mathbf{X}_{k,0}$ and the three anchor points $\mathbf{X}_i$, $\mathbf{X}_j$, and $\mathbf{X}_k$, respectively, near zero.

Absolute Reference Points. We first consider the case of absolute reference points. Since in this case the protein degrees are not affected, one just has to show that the partition function of the restrained ligand Z_L^r can be factored as

$$Z_L^r = Z_r^{BRA,abs} \times \tilde{Z}_L = V_{eff} \times z_\theta \times z_\chi \times \tilde{Z}_L \qquad (37)$$

where V_{eff}, z_θ, and z_χ are given by eqs B4, B9, and B14 of ref 28, respectively. We first transform the Cartesian coordinates $\{\mathbf{r}_1 ... \mathbf{r}_{N_L}\}$ of the ligand atoms to $\{\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k, \mathbf{r'}_1 ... \mathbf{r'}_{N_L-3}\}$; the Jacobian determinant $|\boldsymbol{J}_1|$ is unity or some number. (Consider the following coordinate transformations for a four-atomic ligand: Let $\mathbf{X}_i$ be the COM of the four atoms, $\mathbf{X}_j = \mathbf{r}_2$, and $\mathbf{X}_k = \mathbf{r}_3$. The Jacobian determinant is $(\sum_{i=1}^4 m_i/m_1)^3$.) If atom positions are chosen as the anchor points, then this step obviously is redundant. Then, $\mathbf{X}_j$ and $\mathbf{X}_k$ are replaced by $\mathbf{X}_{ij} = \mathbf{r}$ and $\mathbf{X}_{jk}$; no Jacobian arises. At this point one can separate the integration in $\mathbf{X}_i$, $\int d\mathbf{X}_i \exp(-\beta K_x/2(\mathbf{X}_i - \mathbf{X}_{i,0})^2) = V_{eff}$, from the remaining configurational integral in $\{\mathbf{X}_{ij}, \mathbf{X}_{jk}, \mathbf{r'}_1 ... \mathbf{r'}_{N_L-3}\}$. Next, the new coordinate $\mathbf{X}_{ij} = \mathbf{r}$ is split into a unit vector $\mathbf{e}_{ij}$ and length r; this is equivalent to a transformation into polar coordinates $\mathbf{r} = (r, \theta, \phi)$ with the Jacobian determinant $r^2 \sin\theta$. While the two angles are external DOFs, the length r is an internal DOF, which (together with the Jacobian factor r^2) remains part of $\tilde{Z}_L$. Integrating over the two angles to which the restraint is applied ($\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta \exp[-\beta K_\theta/2(1 - \cos\theta)]$) gives z_θ (eq B9 of ref 28). Finally, the coordinate $\mathbf{X}_{jk}$ is also transformed to polar coordinates. The length and azimuthal angle are internal coordinates and, thus, part of $\tilde{Z}_L$ (which also absorbs the Jacobian determinant of the transformation). The remaining longitudinal angle, which HW denote as χ, is subject to a harmonic restraint; integration over this DOF gives z_χ (eq B14 of ref 28). Overall, one obtains[28]

$$Z_L^r = \left(\frac{2\pi kT}{K_x}\right)^{3/2} \frac{4\pi kT}{K_\theta}(1 - \exp(-K_\theta/kT))\left(\frac{2\pi kT}{K_\chi}\right)^{1/2} \times \tilde{Z}_L = Z_r^{BRA,abs} \times \tilde{Z}_L \qquad (38)$$

The configurational integral that remains after the coordinate transformations described above (including all Jacobian determinants) is $\tilde{Z}_L$ (cf. eq 7), since (as in the VBA) the six DOFs that have been separated from Z_L^r are the external DOFs of the ligand.

Relative Reference Points. In the most general case of restraints relative to the protein, BRA requires three reference points $\mathbf{X}_{i,0}$, $\mathbf{X}_{j,0}$, and $\mathbf{X}_{k,0}$ in addition to the three anchor points in the ligand. These reference points, and consequently the two reference unit vectors $\mathbf{e}_\theta$ and $\mathbf{e}_\chi$ derived from them, are therefore functions of the protein coordinates. In other words, $U_r^{BRA,rel} = U_r^{BRA,rel}(\mathbf{X}_{i,0}(\mathbf{r}_P), \mathbf{e}_\theta(\mathbf{r}_P), \mathbf{e}_\chi(\mathbf{r}_P), \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k)$. Thus, the BRA involves the protein DOFs much more directly than the VBA, and one has to investigate how a factorization of the partition function into protein, ligand, and restraint contributions as in eq 11 can be obtained. To accomplish a separation of variables, one has to choose $\mathbf{X}_{i,0}(\mathbf{r}_P)$, $\mathbf{e}_\theta(\mathbf{r}_P)$, and $\mathbf{e}_\chi(\mathbf{r}_P)$ as the external variables of the protein, taking into account that only the longitudinal polar angle χ of $\mathbf{e}_\chi(\mathbf{r}_P)$ is an external DOF. The same series of coordinate transformations as described in the previous subsection for the ligand needs to be carried out first for the protein. This separation of internal and external protein DOFs generates the nonredundant six reference values required for the BRA algorithm. For the ligand itself one proceeds as in the case of absolute restraints. The only difference is the trivial variable transformation $\Delta\mathbf{X}_i = \mathbf{X}_i - \mathbf{X}_{i,0}$ for the position restraint; then the integration over these three DOFs has been decoupled from the protein. Thus, eq 38 is proven to hold for the case of relative reference points, as well.

Supporting Information Available: Explicit expressions for the translational, rotational, and vibrational free energy contributions. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.

(2) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430−437.
(3) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461−469.
(4) Straatsma, T. P. Free Energy by Molecular Simulation. In *Reviews in Computational Chemistry*, Vol. 9; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1996.
(5) Henchman, R. H.; Essex, J. W. *J. Comput. Chem.* **1998**, *20*, 499−510.
(6) Miller, J. L.; Kollman, P. A. *J. Phys. Chem.* **1996**, *100*, 8587−8594.
(7) Lau, F. T. K.; Karplus, M. *J. Mol. Biol.* **1994**, *236*, 1049−1066.
(8) Lamb, M. L.; Jorgensen, W. L. *J. Med. Chem.* **1998**, *41*, 3928−3939.
(9) Archontis, G.; Simonson, T.; Karplus, M. *J. Mol. Biol.* **2001**, *306*, 307−327.
(10) Koichi, T.; Shimizu, K. *J. Phys. Chem. B* **1998**, *102*, 6419−6424.
(11) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281−283.
(12) Boyer, P. D. *Annu. Rev. Biochem*. **1997**, *66*, 717−749.
(13) Abrahams, J. P.; Leslie, A. G. W.; Lutter, R.; Walker, J. E. *Nature* **1994**, *370*, 621−648.
(14) Yang, W.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 874−879.
(15) Pranata, J.; Jorgensen, W. L. *Tetrahedron* **1991**, *47*, 2491−2501.
(16) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742−3746.
(17) Merz, K. M., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 406−411.
(18) Miyamoto, S.; Kollman, P. A. *J. Am. Chem. Soc.* **1992**, *114*, 3668−3674.
(19) Sneddon, S. F.; Tobias, D. J.; Brooks, C. L., III. *J. Mol. Biol.* **1989**, *209*, 817−820.
(20) Miyamoto, S.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 226−245.
(21) Miyamoto, S.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402−8406.
(22) Helms, V.; Wade, R. C. *Biophys. J.* **1995**, *69*, 810−824.
(23) Sun, Y.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 3599−3604.
(24) Mordasini Denti, T. Z.; van Gunsteren, W. F.; Diederich, F. *J. Am. Chem. Soc.* **1996**, *118*, 6044−6051.
(25) Dixit, S. B.; Chipot, C. *J. Phys. Chem. A* **2001**, *105*, 9795−9799.
(26) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670−681.
(27) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(28) Hermanns, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707−2714.
(29) Göpel, W.; Wiemhöfer, H.-D. *Statistische Thermodynamik*; Spektrum: Heidelberg, Berlin, 2000.
(30) Boresch, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 5145−5154.
(31) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 103−118.
(32) Hermans, J.; Shankar, S. *Isr. J. Chem.* **1986**, *27*, 225−227.
(33) Zhang, L.; Hermans, J. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 433−438.
(34) Mann, G.; Hermans, J. *J. Mol. Biol.* **2000**, *302*, 979−989.
(35) Hermans, J.; Yun, R. H.; Anderson, A. G. *J. Comput. Chem.* **1992**, *13*, 429−442.
(36) Helms, V.; Wade, R. C. *Proteins: Struct., Funct., Genet.* **1998**, *32*, 381−396.
(37) Morton, A.; Baase, W. A.; Matthews, B. W. *Biochemistry* **1995**, *34*, 8564−8575.
(38) Fersht, A. R.; Jakes, R. *Biochemistry* **1975**, *14*, 3350−3356.
(39) Wells, T. N. C.; Knill-Jones, J. W.; Gray, T. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 5151−5156.
(40) Doty, P.; Myers, G. *Discuss. Faraday Soc.* **1953**, *13*, 51−58.
(41) Steinberg, I. Z.; Scheraga, H. A. *J. Biol. Chem.* **1963**, *238*, 172−181.
(42) Searle, M. S.; Williams, D. H. *J. Am. Chem. Soc.* **1992**, *114*, 10690−10697.
(43) Horton, N.; Lewis, M. *Prot. Sci.* **1992**, *1*, 169−181.
(44) Spolar, R. S.; Record, M. T., Jr. *Science* **1994**, *263*, 777−784.
(45) Tidor, B.; Karplus, M. *J. Mol. Biol.* **1994**, *238*, 405−414.
(46) Murphy, K. P.; Xie, D.; Thompson, K. S.; Amzel, L. M.; Freire, E. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 63−67.
(47) Janin, J. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 30−39.
(48) Holtzer, A. *Biopolymers* **1995**, *35*, 595−602.
(49) Baginski, M.; Fogolari, F.; Briggs, J. M. *J. Mol. Biol.* **1997**, *274*, 253−267.
(50) Amzel, L. M. *Proteins: Struct., Funct., Genet.* **1997**, *28*, 144−149.
(51) Ben-Tal, B.; Honig, B.; Bagdassarian, C. K.; Ben-Shaul, A. *Biophys. J.* **2000**, *79*, 1180−118.
(52) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.
(53) Torrie, G. M.; Valleau, J. P. *Statistical Mechanics. Part A: Equilibrium Techniques*; Plenum Press: New York, London, 1977.
(54) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. In *Advances in Chemical Physics*, Vol. LXXI; Wiley: New York, 1988.
(55) DeVoe, H. Theory of the Conformations of Biological Macromolecules in Solution. In *Structure and Stability of Biological Macromolecules*; Timascheff, S. N., Fasman, G. D., Eds.; Dekker: New York, 1969.
(56) Gō, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535−542.
(57) Herschbach, D. R.; Johnston, H. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 1652−1661.
(58) Knox, J. H. *Molecular Thermodynamics*; John Wiley & Sons: New York, 1978.
(59) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(60) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(61) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(62) Shobana, S.; Roux, B.; Andersen, O. S. *J. Phys. Chem. B* **2000**, *104*, 5179−5190.
(63) Boresch, S. *Mol. Simul.* **2002**, *28*, 13−37.
(64) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 119−136.
(65) Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum: New York, 1992.
(66) Minton, A. P. *Curr. Opin. Struct. Biol.* **2000**, *10*, 34−39.
(67) Ellis, R. J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 114−119.
(68) Moore, W. J. *Physical Chemistry*, 5 ed.; Longman: London, 1972.
(69) Janin, J. *Proteins: Struct., Funct., Genet.* **1996**, *24*, i−ii.
(70) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(71) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(72) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133−152.
(73) Brooks, B. R.; Janežič, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522−1542.
(74) Brooks, C. L., III; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312−6325.
(75) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025−9031.
(76) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(77) Hoover, W. G. *Phys. Rev.* **1985**, *A31*, 1695−1697.
(78) Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem. A* **1997**, *101*, 1609−1618.
(79) Kolossváry, I. *J. Phys. Chem. A* **1997**, *101*, 9900−9905.
(80) Tamura, A.; Privalov, P. L. *J. Mol. Biol.* **1997**, *273*, 1048−1060.
(81) Karplus, M.; Janin, J. *Prot. Eng.* **1999**, *12*, 185−186.
(82) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946−950.