JCTC Journal of Chemical Theory and Computation

Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change

David L. Mobley,[†] John D. Chodera,[‡] and Ken A. Dill*,[†]

Department of Pharmaceutical Chemistry, and Graduate Group in Biophysics, University of California at San Francisco, San Francisco, California 94158-2517

Received February 5, 2007

Abstract: Free energy calculations are increasingly being used to estimate absolute and relative binding free energies of ligands to proteins. However, computed free energies often appear to depend on the initial protein conformation, indicating incomplete sampling. This is especially true when proteins can change conformation on ligand binding, as free energies associated with these conformational changes are either ignored or assumed to be included by virtue of the sampling performed in the calculation. Here, we show that, in a model protein system (a designed binding site in T4 lysozyme), conformational changes can make a difference of several kcal/mol in computed binding free energies and that they are neglected in computed binding free energies if the system remains kinetically trapped in a particular metastable state on simulation timescales. We introduce a general "confine-and-release" framework for free energy calculations that accounts for these free energies of conformational change. We illustrate its use in this model system by demonstrating that an umbrella sampling protocol can obtain converged binding free energies that are independent of the starting protein structure and include these conformational change free energies.

INTRODUCTION

Computational tools are becoming increasingly important in drug discovery.[1] A major goal is to use these methods to predict (absolute or relative) protein−ligand binding free energies. A great deal of effort[2−4] has been focused on identifying which protein structures (i.e. *apo*, *holo*, or optimized in some manner) work best for estimating binding affinities. This emphasis on a single bound structure or conformation begs the question, "Can protein−ligand binding free energies be accurately predicted only a single protein conformation, or only some of the relevant protein conformations are considered?". We demonstrate here that the answer is a decisive *no* in at least the model system considered here. There can be significant strain energies and free energy costs associated with trapping a protein into *any* metastable state, and, as we show here, the neglect of these costs can lead to substantial errors that depend on the metastable state chosen. (Here, we will use the term "structure" to refer to a single static structure and the term "metastable state" to refer to a favorable region of configuration space (set of structures) that is kinetically distinct from other such regions).

Computed binding free energies are often sensitive to the starting protein structure, even with alchemical free energy methods,[5−10] which should not be the case if these simulations are converged. We believe this is for a similar reason: Even if full protein flexibility is allowed, the full range of relevant protein states may not be accessible on simulation timescales. This means that the protein is kinetically trapped in a particular metastable state, and the free energy cost of this trapping is neglected. Here, the problem is fundamentally a kinetic one: Large energy barriers can separate metastable protein states and trap the protein in a single metastable state on simulation timescales. Unfortunately, this trapping is inevitable whenever energy barriers are sufficiently large,[11] yet inadequate sampling even at the level of a single side chain rotameric state can lead to a difference in several kcal/mol in computed binding free energies.[8] The problem is that it is necessary to adequately sample multiple relevant protein metastable states, including at least the metastable states containing both the *apo* and *holo* structures.

Here, we describe a framework we call "confine-and-release" for computing absolute binding free energies that correctly accounts for multiple relevant metastable states, such as protein conformational changes on ligand binding. The framework is general, in that it may be implemented in a number of different ways. We demonstrate the framework in a model binding site using one particular approach based on umbrella sampling, below.

In this work, we will refer to the problem of kinetic trapping or confinement as "[virtual] confinement", to distinguish it from real confinement, where an external biasing potential is used. The confine-and-release approach

* Corresponding author e-mail: dill@maxwell.compbio.ucsf.edu.
[†] Department of Pharmaceutical Chemistry.
[‡] Graduate Group in Biophysics.

10.1021/ct700032n CCC: $37.00
Published on Web 05/23/2007

discussed here can deal with both cases, but we illustrate it here with virtual confinement.

The basic theory underlying absolute binding free energy calculations has previously been described in detail (for example, in refs 12 and 13). The absolute binding free energy is given as

$$\Delta G_{bind} = -k_B T \ln \frac{C^o}{8\pi^2} \frac{\sigma_P \sigma_L}{\sigma_{PL}} \frac{Z_{PL}}{Z_P Z_L} + P^o \Delta V_{PL}$$

where the protein-complex partition function is given by

$$Z_{PL} = \int_{complex} e^{-\beta U(\vec{r})} d\vec{r}$$

which is an integral over all of the protein−ligand conformations defining the bound state, and Z_P and Z_L are the corresponding partition functions consisting of integrals for the protein and ligand alone in solvent, respectively. C° denotes the standard concentration (1 M), and the σ factors are the symmetry numbers for the protein, for the ligand, and for the complex. These terms as well as the $P^o \Delta V_{PL}$ pressure-volume work term relate to the standard state and are explained in detail elsewhere.[14]

The essential point here is that evaluating the binding free energy necessarily involves integrating over all of the relevant (low potential energy) conformations of the protein and ligand, including all metastable states. If that integration is incomplete, as in the case of inadequate sampling, the quantity calculated will not be a true binding free energy. In such cases of kinetic trapping, the free energy that is calculated can be called a "confined" binding free energy—it measures the binding free energy of the system [virtually] confined to a metastable state (for example, the region corresponding to the *holo* structure) and hence neglects certain components of the true binding free energy such as strain energies. This observation is related to that made earlier by Straatsma and McCammon in the context of solvation for molecules with multiple relevant rotameric states: Unless all relevant metastable states are sampled in some manner, computed free energies are "unreasonable" and incorrect.[15]

We illustrate the problem with the example of *p*-xylene binding in a simple apolar cavity (an engineered cavity in T4 lysozyme) studied computationally by Deng and Roux.[8] Here, a single valine side chain reorients upon ligand binding (as seen by comparing the *apo* and *holo* structures[19]).

We use simulation protocols employed previously[14] with minor modifications described in the Supporting Information. These modifications involve improved parameters for the Particle mesh Ewald[16] treatment of long-range electrostatics, addition of a separate vacuum calculation in order to finish the cycle for computing binding free energies, and addition of a long-range correction term to account for attractive dispersion interactions between the ligand and protein that are neglected when simulations are run with a short cutoff. Very briefly, the overall procedure involves first restraining the ligand in complex, then annihilating the ligand's electrostatic interactions, followed by decoupling its Lennard-Jones interactions. The restraints are then analytically removed, and this is equivalent to having a protein with no ligand, plus a noninteracting, neutral ligand in solvent. The ligand electrostatic interactions are then restored in solvent, completing the thermodynamic cycle. The free energy of making each of these transformations is computed using free energy methods with a series of separate simulations at different alchemical intermediate states (λ values).

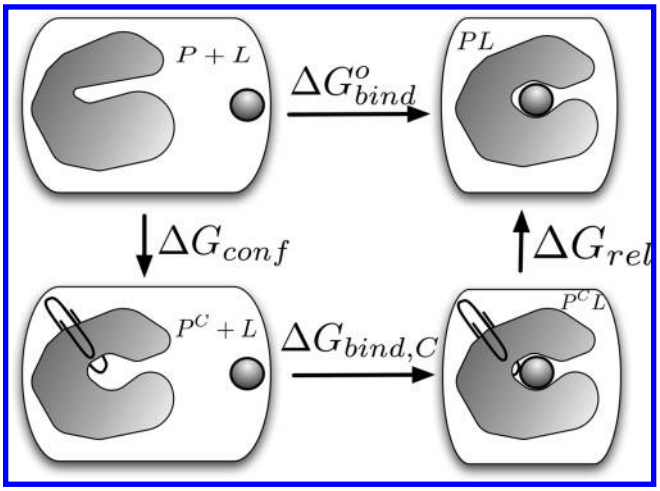

Figure 1. Thermodynamic cycle for the confine-and-release framework. The quantity we want to calculate is ΔG^o_{bind} (top), the free energy difference for the process P + L → PL. Kinetic trapping (virtual confinement) or deliberate confinement can keep conformational changes from being sampled (shown graphically by a paperclip). When this happens, computed free energies are actually confined binding free energies, $\Delta G^o_{bind,C}$ (bottom arrow). To relate these to true binding free energies, it is necessary to compute the free energy of confining the protein in the absence of the ligand (left arrow) and releasing the protein in the presence of the ligand (right arrow).

We start from the *apo* structure. We observe that the system remains trapped in the metastable state containing that structure over the course of all equilibration and production trajectories involved in the free energy calculation (1.11 ns at each λ value). The resulting computed binding free energy (at standard pressure and 300 K) is -2.96 ± 0.06 kcal/mol (where the uncertainty represents 1 SD over a set of block bootstrap trials as described in the Supporting Information and previously[14]). If, instead, we start from the *holo* structure, we compute a binding free energy of -7.27 ± 0.09 kcal/mol. If we examine the valine side chain χ_1 dihedral angle as a function of time for every simulation in these free energy calculations, we find that, in each case, it remains in its initial rotameric state. The valine does not cross its torsional energy barrier on simulation timescales. This causes significant errors: One computed binding free energy indicates *p*-xylene is a millimolar binder; the other indicates it is a micromolar binder.

We solve this problem using the "confine-and-release" framework, depicted in the thermodynamic cycle in Figure 1; there, confinement (in this case virtual confinement) is illustrated by a paper clip. We begin by recognizing that our calculated free energies are "confined" binding free energies, that is, free energies for binding of the ligand to a protein that is restricted to a particular metastable state. Then, to compute the true binding free energy, we must add the free energy of confining the protein to that metastable state when no ligand is bound and the free energy of releasing the protein from its confinement when the ligand is bound. Hence $\Delta G^o_{bind} = \Delta G_{conf} + \Delta G^o_{bind,C} + \Delta G_{rel}$. In this expression, ΔG^o_{bind} is the true (standard) binding free energy; $\Delta G^o_{bind,C}$ is the standard binding free energy of the ligand to the confined protein; ΔG_{conf} is the free energy of confining the protein to

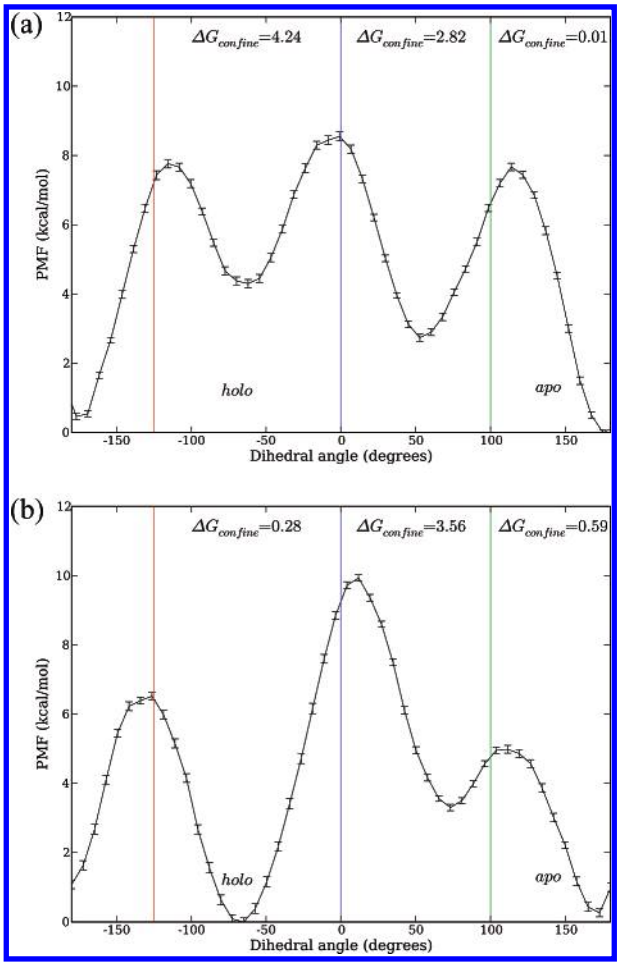

Figure 2. Potential of mean force for rotating the valine 111 side chain, with (b) and without (a) the ligand. Above each of the three regions is shown the free energy of confining Val111 to that metastable state. The *apo* metastable state corresponds to the first region on the left and the far right region (since the dihedral angle is periodic). Error bars represent statistical uncertainties corresponding to 1 SD. Uncertainties for confinement to each well are given in the text.

this smaller region of configuration space in the unbound state; and ΔG_{rel} is the free energy of releasing the protein from conformational confinement in the bound state. This can be thought of as a generalization of conformational biasing potentials.[8]

Free energies of confinement and release can be computed using a variety of different algorithms. Here, since there is a single relevant degree of freedom that needs to be sampled, we employed umbrella sampling.[17] We computed the potential of mean force (PMF) for rotating the side chain of Val111 throughout its range of motion in both the bound and unbound states (Figure 2) (details available in the Supporting Information). From the PMF, we computed the free energy of confining the side chain to each rotameric state (as described in the Supporting Information). To test reproducibility of the corrected true free energy, the entire confine-and-release procedure was performed twice: Once using the *apo* structure and the associated metastable state (beginning from the *apo* crystal structure) for the binding calculation, and once using the *holo* structure (and metastable state) for the binding calculation. The same framework applies in either case.

Using the *apo* metastable state, we compute a confinement free energy in the unbound state of 0.01 ± 0.04 and a release free energy of -0.6 ± 0.1 kcal/mol in the bound state. Combining this with the computed confined binding free energy of -2.96 ± 0.06 kcal/mol yields a total binding free energy of -3.5 ± 0.2 kcal/mol. Alternatively, using the *holo* metastable state, the confinement free energy is 4.2 ± 0.2 kcal/mol, and the release free energy is 0.28 ± 0.08 kcal/mol, which, when added to the computed confined binding free energy of -7.27 ± 0.09 kcal/mol, yields a total binding free energy of -3.3 ± 0.2 kcal/mol. The difference between the total binding free energies computed from the different crystal structures is now only 0.2 ± 0.3 kcal/mol—statistically indistinguishable from zero. Hence, we believe these values now represent the overall binding free energy, corrected for inadequate sampling of Val111. In this case, the experimental binding free energy is -4.67 ± 0.06 kcal/mol, so our approach substantially improves agreement with experiment, especially when beginning from the *holo* structure.

Figure 2 shows that, for *p*-xylene, a single rotameric state dominates when the ligand is absent (Figure 2a), and a different rotameric state dominates when the ligand is present (Figure 2b), although in (Figure 2b), both rotameric states are relevant—that is, both states contribute significant fractions to the free energy. In general, the relevant rotameric state may differ in the presence and absence of the ligand, or there may be multiple relevant states in either case.

Previous work on this binding site, beginning from the *holo* structure for each ligand, produced binding free energies that were 2.05 to 4.40 kcal/mol too negative relative to experiment[8] for those compounds where Val111 reorients on ligand binding (*p*-xylene, *o*-xylene, and *n*-butylbenzene, isobutylbenzene).[19] Indeed, these compounds were essentially the worst outliers in that study. Here, due to kinetic trapping, we had to apply a positive correction of 3.9 kcal/mol for *p*-xylene beginning from the *holo* structure. Though the previous work used a different force field and parameters, it seems likely that kinetic trapping of Val111 can explain a significant portion of the observed errors there. For example, if we applied our correction to their calculated value for *p*-xylene (−9.06 kcal/mol), the resulting binding free energy would be −5.06 kcal/mol (calculated) versus −4.67 kcal/mol (experiment).

Here, the confine-and-release technique was applied to a single degree of freedom. As other situations will undoubtedly require careful sampling of more than a single (known) degree of freedom, the calculation of free energies of confinement and release from potentials of mean force is not necessarily a general strategy for applying this framework. Rather, the key points here are as follows: First, correct binding free energies can only be obtained when protein conformational change is correctly accounted for. Second, protein conformational change contributes substantially to the overall binding free energy, even for changes as small as the reorientation of single side chains. Thus, protein conformational changes should not simply be ignored in binding free energy calculations.

To compute confine-and-release free energies with the umbrella sampling approach discussed here, there are several requirements. First, one must know (i.e., crystallographically) or be able to predict (i.e., from side chain Monte Carlo sampling[4]) all of the relevant slow degrees of freedom.

Second, there must be relatively few of these degrees of freedom so that deliberate sampling of them is tractable. In this particular binding site, crystallographic evidence suggests that the only side chain reorientation on ligand binding is that of Val111,[19] thus it is straightforward to apply this umbrella sampling approach. In general, however, umbrella sampling may prove impractical. But the confine-and-release approach itself (Figure 1) only requires a method of computing the confine and release free energies; this need not be done with umbrella sampling.

The confine-and-release cycle used here, involving confining and releasing the protein to compute true binding free energies, can easily be extended to a variety of other applications. In the example above, [virtual] confinement is due to kinetic trapping. But deliberate confinement by external restraints may also be useful. This could help, for example, for proteins that undergo relatively large conformational changes on ligand binding, such flap closure in HIV protease. Without this confinement, the protein could begin to deform back to its *apo* structure as the ligand is alchemically removed, leading to sampling problems. These sampling problems can be severe: At some alchemical intermediate states, *both* metastable states could be relevant, and the protein would need to sample both several times during the simulation. In HIV protease, for example, these conformational changes may take place on the microsecond to millisecond time scale and are difficult to sample even with long molecular dynamics trajectories.[18] Thus, this confinement approach can also potentially aid convergence at intermediate alchemical states.

We conclude that computing binding free energies requires more than just computing the binding free energy of the ligand to a particular conformational state of the protein; it also requires a calculation of the free energy associated with confining the protein to that particular conformational state with and without the ligand present. These confinement free energies can be substantial, even for the relatively rigid binding site considered here. Elsewhere, we have noted that similar problems can arise when sampling ligand orientations.[14] Unless free energy calculations include sufficient sampling to adequately include these conformational changes at all stages of the transformation, computed "binding free energies" are not true binding free energies. In short, a dependence of free energy estimates on initial protein or ligand structure can indicate that simulations are not adequately sampling the relevant regions of configuration space. The confine-and-release framework we introduce here can be used to design approaches that isolate and solve these sampling problems in a systematic and controlled manner for free energy calculations.

The importance of conformational change in binding free energies has ramifications that extend beyond just alchemical free energy calculations. Virtual screening methods that rely on docking and scoring using a single structure need to reconsider the assumption that binding free energies can be estimated given an appropriate bound structure. Free energy costs associated with trapping the protein to the *holo* structure, or to any structure chosen, may be significant and probably need to be correctly accounted for to accurately predict binding free energies. This problem cannot be avoided simply by comparing relative binding free energies of different ligands, either. In this binding site, for example, it is known that some ligands bind *without* reorientation of the Val111 side chain, while others require the reorientation seen here in the case of *p*-xylene.[19] This means that free energy costs required to bind different ligands can be substantially different—up to several kcal/mol, based on the data presented here. Thus, when estimating relative binding free energies using the same protein structure, errors will be different for different ligands rather than canceling out.

In summary, the confine-and-release framework presented here provides a rigorous way to correct for inadequate or restricted computational sampling of protein degrees of freedom in ligand binding free energy calculations. This approach can give binding free energies that are independent of the starting protein structure (i.e., *apo* or *holo*) and therefore yield true binding free energies for the given the force field. Here we have demonstrated this approach using an umbrella sampling technique for computing the confine-and-release free energies; sampling requirements will probably limit this particular technique to accounting for inadequate sampling of a limited number of degrees of freedom. But the framework is more general.

ACKNOWLEDGMENT

We thank Benoît Roux (University of Chicago) for a critical reading of the manuscript. J.D.C. was supported in part by HHMI and IBM predoctoral fellowships. D.L.M. and K.A.D. acknowledge NIH grant GM63592, Anteon Corporation grant USAF-5408-04-SC-0008, and a UCSF Sandler Award.

Supporting Information Available: Details of umbrella sampling calculations, simulation protocols, and computation of confinement and release free energies. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

(1) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813−1818.

(2) McGovern, S. L.; Shoichet, B. K. Information Decay in Molecular Docking Screens Against Holo, Apo, and Modeled Conformations of Enzymes. *J. Med. Chem.* **2003**, *46*, 2895−2907.

(3) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Molecular Mechanics Methods for Predicting Protein-Ligand Binding. *J. Chem. Inf. Model*. **2006**, *46*, 243−253.

(4) Meiler, J.; Baker, D. ROSETTALIGAND: Protein-Small Molecule Docking with Full Side-Chain Flexibility. *Proteins: Struct., Funct., Bioinf*. **2006**, 66, 538−548.

(5) Wang, J.; Deng, Y.; Roux, B. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophys. J.* **2006**, *91*, 2798−2814.

(6) Shirts, M. R. Calculating Precise and Accurate Free Energies in Biomolecular Simulations. Ph.D. Dissertation, Stanford University, 2004. ProQuest location: http://wwwlib.umi.com/dissertations/fullcit/3153076 (accessed Feb 1, 2007).

(7) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. Direct Calculation of the Binding Free Energies of FKBP Ligands. *J. Chem. Phys.* **2005**, *123*, 084108.

(8) Deng, Y.; Roux, B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *J. Chem. Theor. Comput.* **2006**, *2*, 1255−1273.

(9) van den Bosch, M.; Swart, M.; Snijderst, J. G.; Berendsen, H, J. C.; Mark, A. E.; Oostenbrink, C.; van Gunsteren, W. F.; Canters, G. W. Calculation of the Redox Potential of the Protein Azurin and Some Mutants. *Chem. Bio. Chem.* **2005**, *6*, 738−746.

(10) Zhou, Y.; Oostenbrink, C.; Jongejan, A.; Hagen, W. R.; de Leeuw, S. W.; Jongejan, J. A. Computational Study of Ground-State Chiral Induction in Small Peptides: Comparison of the Relative Stability of

Selected Amino Acid Dimers and Oligomers in Homochiral and Heterochiral Combinations. *J. Comput. Chem.* **2006**, *27*, 857−867.

(11) Leitgeb, M.; Schroder, C.; Boresch, S. Alchemical Free Energy Calculations and Multiple Conformational Substates. *J. Chem. Phys.* **2005**, *122*, 084109.

(12) Gilson, M. K.; Given, J. A.; Bush, G. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047−1069.

(13) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.

(14) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the Use of Orientational Restraints and Symmetry Number Corrections in Alchemical Free Energy Calculations. *J. Chem. Phys.* **2006**, *125*, 084902.

(15) Straatsma, T. P.; McCammon, J. A. Treatment of Rotational Isomers in Free Energy Evaluations. Analysis of the Evaluation of Free Energy Differences by Molecular Dynamiccs Simulations of Systems with Rotational Isomeric States. *J. Chem. Phys.* **1989**, *90*, 3300−3304.

(16) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, Lee, G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(17) Kumar, S. J.; Rosenberg, M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analaysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(18) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. HIV-1 Protease Flaps Spontaneously Open and Reclose in Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915−920.

(19) Morton, A.; Matthews, B. W. Specificity of Ligand Binding in a Buried Nonpolar Cavity of T4 Lysozyme: Linkage of Dynamics and Structural Plasticity. *Biochemistry* **1995**, *34*, 8576−8588.

CT700032N