Computational Alchemy To Calculate Absolute Protein−Ligand Binding Free Energy

Volkhard Helms[†] and Rebecca C. Wade*

Contribution from the European Molecular Biology Laboratory, 69012 Heidelberg, Germany

Received November 10, 1997

Abstract: The ability to reliably compute accurate protein−ligand binding affinities is crucial to understanding protein−ligand recognition and to structure-based drug design. A ligand's binding affinity is specified by its absolute binding free energy, ΔG_{bind}, the free energy difference between the bound and unbound states. To compute accurate free energy differences by free energy perturbation (FEP), "alchemical" rather than physical processes are usually simulated by molecular dynamics simulations so as to minimize the perturbation to the system. Here, we report a novel "alchemistic" application of the FEP methodology involving a large perturbation. By mutating a ligand with 11 non-hydrogen atoms into six water molecules in the binding site of a protein, we computed a ΔG_{bind} within 3 kJ/mol of the experimental value. This is the first successful example of the computation of ΔG_{bind} for a protein:ligand pair with full treatment of the solvent degrees of freedom.

Introduction

Fundamental to understanding how proteins recognize their ligands and to the rational design of proteins and drugs is the ability to compute protein−ligand binding affinities. While a number of approaches are being pursued toward computing ligand−protein binding affinities with use of empirical models,[1−4] and models that treat parts of the system as a continuum,[5,6] molecular dynamics simulations in full atomic detail employed with the FEP methodology offer the prospect of a generally applicable rigorous "first principles" solution to the "binding problem". In applications to protein−ligand binding, the FEP methodology has usually been used to compute $\Delta\Delta G$s, i.e., differences between the binding free energies of two similar ligands to one protein target, or of one ligand to a protein and a mutant. The "alchemical" mutations involved in such simulations are often restricted to mutation of a single non-hydrogen atom although mutations of several non-hydrogen atoms have been performed.[7,8] In the few previous examples of computing absolute ligand binding free energies, ΔG_{bind}, to proteins, either no solvent was present in the binding site[9] or full relaxation of the solvent in the binding site at all stages of the simulation was not permitted,[10−13] possibly due to limitations of computational resources. Although qualitative agreement with experiment could be obtained in some of these calculations (with differences between calculated and experimental values of ∼12 kJ/mol), a procedure with full treatment of the important degrees of freedom of the solvent is necessary to reliably obtain accurate results (within 4 kJ/mol of experimental values). To develop and demonstrate such a procedure, we chose the binding of the substrate, camphor, to cytochrome P450cam from *Pseudonomas putida* as an especially well-suited model system because camphor binds in a buried active site isolated from bulk solvent. This facilitates the identification of those solvent molecules that are expelled from the active site into bulk solvent upon ligand binding, and thus contribute to the thermodynamics of the binding process. In addition, cytochrome P450cam has long served as a model for understanding the structure−function relationships of the cytochrome P450 superfamily of enzymes and has thus been very well-characterized in biophysical experiments.[14] Furthermore, cytochrome P450cam is, itself, an important biotechnological target for bioremediation tasks,[15] and the ability to compute binding constants for ligands to it is of clear value in design projects.

The active site of cytochrome P450cam as observed crystallographically is shown with camphor bound[16] in Figure 1a, and in the unbound state[17] in Figure 1b. In the absence of camphor, one ordered water molecule was observed as a ligand to the heme iron and a second region of electron density in the active

* To whom correspondence should be addressed at the European Molecular Biology Laboratory, Postfach 10.2209, 69012 Heidelberg, Germany, Tel. +49-6221-387553, Fax +49-6221-387517, e-mail: wade@embl-heidelberg.de.

[†] Current address: Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92037.

(1) Böhm, H. J. *J. Comput. Aided Mol. Des.* **1992**, *6*, 593−606.
(2) Åqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng.* **1994**, *7*, 385−391.
(3) Head, R. D.; Smyte, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. *J. Am. Chem. Soc.* **1996**, *118*, 3959−3969.
(4) Paulsen, M. D.; Ornstein, R. L. *Protein Eng.* **1996**, *9*, 567−571.
(5) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161−185.
(6) Froloff, N.; Windemuth, A.; Honig, B. *Protein Sci.* **1997**, *6*, 1293−1301.
(7) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395−2417.
(8) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.
(9) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707−2714.
(10) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Protein Eng.* **1992**, *5*, 215−228.
(11) Miyamoto, S.; Kollman, P. A. *Proteins* **1993**, *16*, 226−245.
(12) Miyamoto, S.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402−8406.
(13) Åqvist, J. *J. Comput. Chem.* **1996**, *17*, 1587−1597.
(14) Poulos, T. L.; Raag, R. *FASEB J.* **1992**, *6*, 674−679.
(15) Kellner, D. G.; Maves, S. A.; Sligar, S. G. *Curr. Opin. Biotechnol.* **1997**, *8*, 274−278.
(16) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *J. Mol. Biol.* **1987**, *195*, 687−700.
(17) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Biochemistry* **1986**, *25*, 5314−5322.

S0002-7863(97)03853-5 CCC: $15.00
Published on Web 03/12/1998

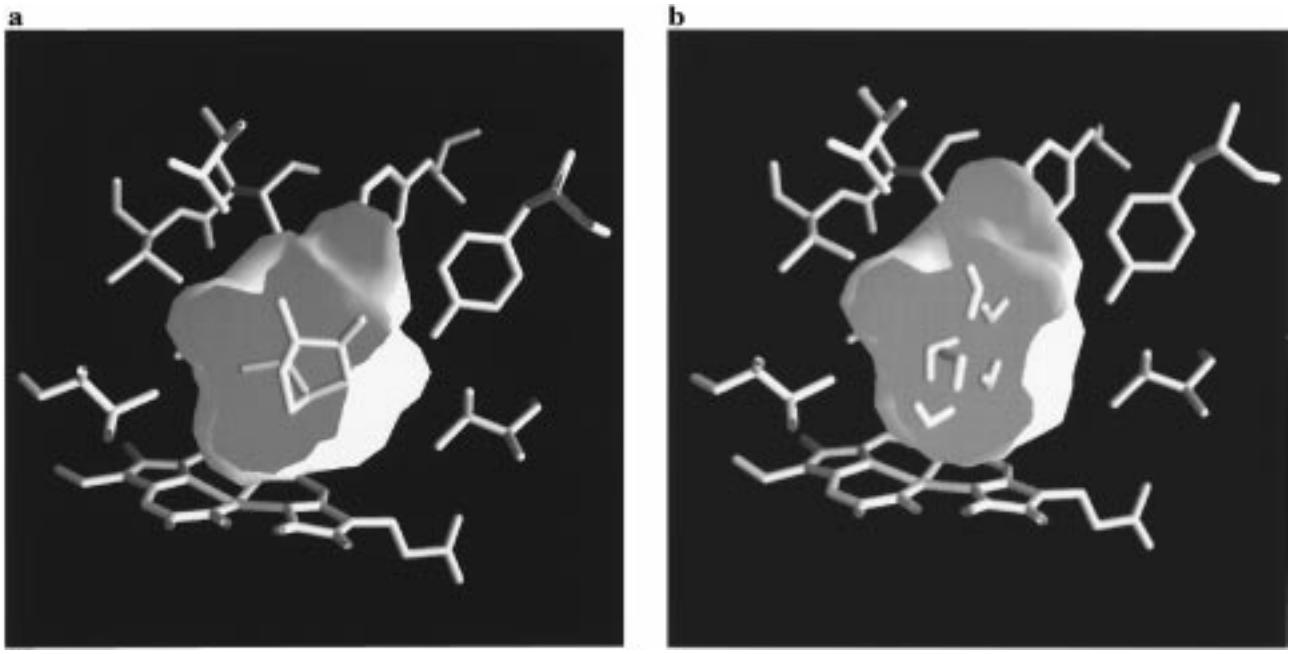

Figure 1. The active site of cytochrome P450cam in the crystal structures with (a) camphor bound[16] and (b) six water molecules bound.[17] The volume of the active site cavity is shown by a molecular surface constructed for the empty active site with a probe of radius 0.14 nm that was generated with the GRASP program.[30]

site was interpreted as corresponding to five additional water molecules forming a network of hydrogen bonds. We have previously calculated the thermodynamically most favorable number of water molecules to occupy the unliganded active site of cytochrome P450cam[18] using the same methodology as employed in this study. In agreement with the crystallographic determination[17] and studies employing hydrostatic pressure,[19] the preferred hydration state was found to be 6 water molecules. In this study, camphor was exchanged against 6 water molecules during molecular dynamics simulations of a region of the protein encompassing the active site and the associated free energy change was computed. Combining this free energy difference with free energy differences for the other components of ΔG_{bind} that we have previously computed (and which correspond to transformations in solution)[20,21] resulted in a computed ΔG_{bind} in excellent agreement with the experimental value.

Materials and Methods

Crystal structures of the substrate-free and camphor-bound cytochrome P450cam from the Brookhaven Protein Databank (identifier codes: 1phc and 2cpp, respectively) were used. Molecular dynamics simulations were performed with the ARGOS program[22] using the GROMOS87 force field[23] with the SPC/E water model.[24]

To compute the ΔG_{bind} for camphor to cytochrome P450cam, the free energy changes on exchanging camphor with six water molecules both in the protein's active site and in solution must be calculated. The free energy change in solution was calculated by us previously from the simulation of two separate transformations: the removal of camphor from water[21] (transformation (2) in Figure 3) and the insertion of a water molecule into bulk water[20] (giving the free energy change for transformation (3) in Figure 3). Simulations in solution were carried out for systems consisting of one solute molecule placed in a periodic

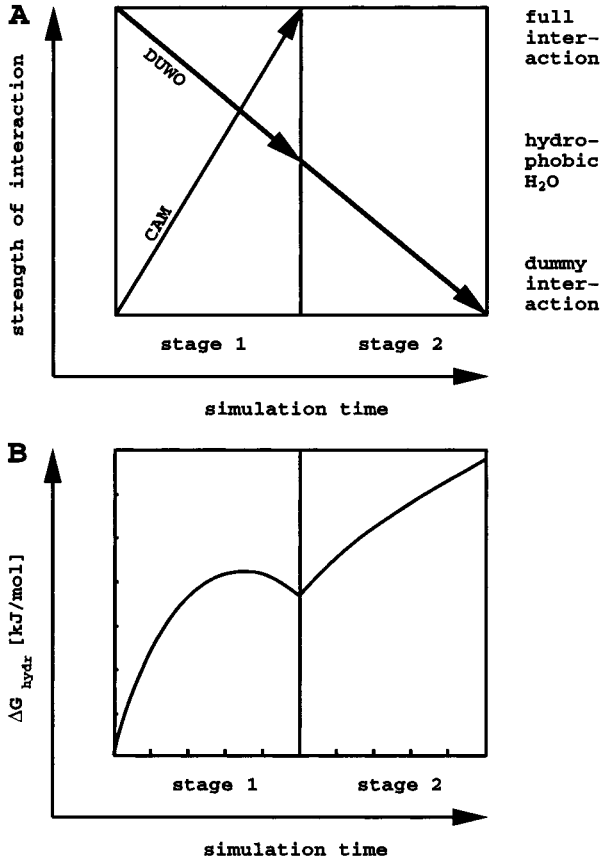

Figure 2. Schematic plots showing (A) how the molecular interactions were modified and (B) how the free energy change accumulated during the molecular dynamics simulations in which six water molecules (DUWO) were transformed into camphor (CAM) in the active site of cytochrome P450cam in two stages. No interactions were calculated between camphor and the active site water molecules. See text for details.

(18) Helms, V.; Wade, R. C. *Proteins* Accepted for publication.
(19) DiPrimo, C.; Hui Bon Hoa, G.; Douzou, P.; Sligar, S. *Eur. J. Biochem.* **1990**, *193*, 383−386.
(20) Helms, V.; Wade, R. C. *Biophys. J.* **1995**, *69*, 810−824.
(21) Helms, V.; Wade, R. C. *J. Comput. Chem.* **1997**, *18*, 449−462.
(22) Straatsma, T. P.; McCammon, J. A. *J. Comput. Chem.* **1990**, *11*, 943−951.
(23) vanGunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation Library* **1987**.
(24) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.

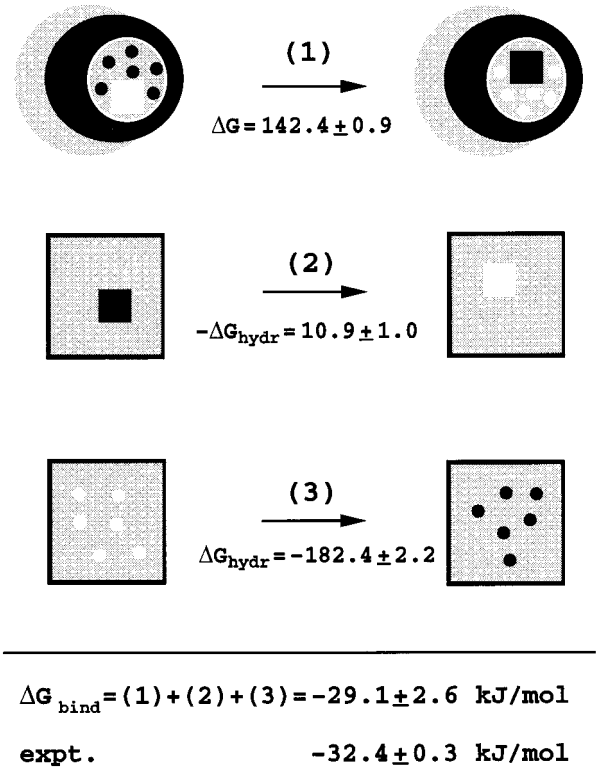

Figure 3. Schematic diagram showing the thermodynamic transformations used to calculate the absolute binding free energy of camphor to cytochrome P450cam. Transformation (1) was simulated in this work, transformation (2) was simulated previously to compute the free energy of hydration of camphor,[21] and transformation (3) is equivalent to 6 times the excess hydration free energy of an SPC/E water molecule without a polarization correction computed previously.[20] The absolute binding free energy of camphor to cytochrome P450cam is the sum of parts 1, 2, and 3. The errors given are calculated from the autocorrelation of the data.[31] The small square represents camphor and the small circles represent water molecules (black: present; white: absent). The system simulated is represented by a square (periodic box of water molecules) or circles (the active site of cytochrome P450cam surrounded by the sphere of solvated protein whose dynamics is simulated and the remaining protein which is held fixed).

box of water molecules. For each system, the solute molecule was gradually removed during simulations in which the free energy change was computed by Multiconfigurational Thermodynamic Integration (MCTI),[25] using separation shifted potential scaling[26] to ensure proper convergence. Full details of these simulations are given in refs 20 and 21. The simulation of the transformation of six water molecules into camphor in the active site of cytochrome P450cam (transformation (1) in Figure 3) was performed as follows.

The parametrization and setup were the same as in our previous work,[20,27] except that the interaction between aliphatic carbon atoms and water oxygens in the GROMOS87 force field[23] was modified as described in our recent study on the computation of hydration free energies for small molecules[21] (the C12 parameter for the repulsive part of the Lennard-Jones interaction was changed from 421 to 690 [kcal Å^{-12} mol^{-1}]$^{1/2}$). Only the dynamics of a spherical region of the protein of radius 15 Å around the geometric center of the active site was simulated; atoms at a distance between 15 and 17 Å from the center were harmonically restrained to their crystallographic positions, and the rest of the solvated protein was held stationary.

Free energy changes were computed in two stages by MCTI[25] during molecular dynamics simulations. As shown in Figure 2a, the intermolecular interactions of camphor were smoothly changed from dummy character (none) to full interactions during stage 1. Simultaneously, the electrostatic interactions of the active site water molecules were reduced from full interactions to those of artificial "hydrophobic water molecules" with partial atomic charges scaled by 0.413. This scaling factor gave the smoothest free energy curves in our previous study.[18] All the remaining intermolecular interactions of the active site water molecules were then removed completely during stage 2 while maintaining the full interactions of camphor. No interactions were calculated between camphor and the active site water molecules, corresponding to the principle of dual topologies.[28] Stages 1 and 2 each consisted of 21 windows with 6000 time steps of equilibration and 45000 time steps of data acquisition. A time step of 2 fs was used, and thus each of the two stages corresponded to 21 × 102 ps or ca. 2.2 ns of simulation. Such long simulation times and the use of separation shifted potential scaling[26] to facilitate the insertion and deletion of particles ensured proper convergence of the free energy changes.

At the beginning of each MCTI window, the coordinates of camphor were assigned from the crystal structure of cytochrome P450cam with camphor bound. Except for the first 13 MCTI windows of stage 1, camphor stayed close to its crystallographic conformation, maintaining the hydrogen bond to Tyr96. The conformational space accessible to all perturbed atoms was reduced to a spherical region around the center of the active site by a flat-bottomed harmonic well potential with a force constant of 1000 kJ nm^{-2} mol^{-1} and a bottom radius of 0.6 nm.[20] The computed free energy change was corrected for the contribution of this restraint potential as described in ref 18. Finally, a correction factor to give the binding free energy for 1 M standard state was also computed as described in ref 18.

Results and Discussion

Figure 2 shows how the particle interactions were modified during the simulations, and the free energy difference that was obtained by this integration path. Note the smooth and nearly monotonic shape of the free energy curve in Figure 2B; these characteristics are important for obtaining converged results. Particularly important in obtaining satisfactory convergence was the choice of perturbation path. The scheme shown in Figure 2A was validated with computations of the energetically most favorable hydration state of the active site.[18] By testing several schemes for the exchange of 6 water molecules with 6 different water molecules in the active site, for which the computed free energy change should be zero, a procedure similar to that shown in Figure 2a was found to give the best results. The amount of simulation time necessary (ca. 100 ps per simulation window) was also estimated from these control simulations for water exchange in the active site. The use of the flat-bottomed harmonic well restraining potential is also essential for convergence, and we tested this previously in computations of the hydration free energy of interfacial cavities containing single water molecules[20] and the whole active site.[18]

The computed free energy change on converting six water molecules into a camphor molecule in the active site of cytochrome P450cam was the sum of 73.6 ± 0.7 kJ/mol for stage 1, 62.4 ± 0.5 kJ/mol for stage 2, the correction for the restraint potential of 1.3 kJ/mol for the first window of stage 1 and 2 × 10^{-3} kJ/mol for the last window of stage 2, and the correction to 1 M standard state of +5.1 kJ/mol, giving a total of 142.4 ± 0.9 kJ/mol.

The results of this work were combined with those from our previous work[20,21] to give the absolute binding free energy of camphor to cytochrome P450cam. The thermodynamic cycle used and the calculated values are shown in Figure 3. The computed value of ΔG_{bind} of −29.1 ± 2.6 kJ/mol is ca. 3 kJ/

(25) Helms, V.; Deprez, E.; Gill, E.; Barret, C.; Hoa, G. H. B.; Wade, R. C. *Biochemistry* **1996**, *35*, 1485−1499.
(26) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *91*, 3631−3637.
(27) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487−1493.
(28) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025−9031.

mol above the experimental value of -32.4 ± 0.3 kJ/mol.[27] The ΔG_{bind} and ΔG_{hydr} computed from our calculations are within 4 kJ/mol of the experimental value, a border that has been termed "chemical accuracy".[29]

This quality of agreement with experiment for ΔG_{bind} can be attributed to (i) the treatment of the contribution of water molecules, (ii) the quality of the force field description, and (iii) a simulation protocol designed to ensure sufficient sampling of the relevant degrees of freedom as discussed above. The explicit consideration of solvent is essential for computing ΔG_{bind}. This was also the case in our previous computations of binding enthalpies for a set of camphor analogues binding to cytochrome P450cam, for which very good agreement between experimental and calculated binding enthalpies was obtained when including all solvent contributions.[27] These calculations were performed with a modification of the Gromos87 force field which gives a better description of carbon−water interactions and whose parameters were derived and validated by computing ΔG_{hydr} for small compounds that are chemically related to camphor.[21]

Although calculations for cytochrome P450cam are facilitated by the buried nature of its active site, meaning that the water molecules displaced on binding are clearly defined, as well as the similarity of the protein structure in the presence and absence of substrate (see Figure 1), these are not absolute requirements. The method presented here should be generally applicable to the calculation of protein−ligand ΔG_{bind} values. Although in the case considered here the ligand displaces all water molecules from the binding site, the approach could be extended to cases in which some of the water molecules in the binding site remain on ligand binding and mediate the protein−ligand interaction. ΔG_{bind} could be calculated with use of the dual topology scheme by simultaneously removing all water molecules from the binding site and introducing the ligand together with the water molecules with which it binds. Alternatively, ΔG_{bind} could be calculated by removing only the number of water molecules displaced from the binding site and simultaneously introducing the ligand. More extensive sampling than in the case studied here may be necessary if the number and location of the water molecules in the ligand-bound state are unknown. Another extension of the approach would be to ligands binding at surface-exposed sites on proteins. In such cases, the same general procedure should be applicable but the effects of bulk solvent on the simulated ligand−water molecule exchange process will need to be examined carefully.

Conclusions

In conclusion, we have described a procedure for computing the absolute binding free energy of a ligand to a protein with full consideration of the contribution of water molecules by free energy perturbation using molecular dynamics simulations. A computed value within 3 kJ/mol of the experimental value is obtained for the binding of camphor to cytochrome P450cam. The excellent agreement provides support for the validity of the model and the calculation procedure which is expected to be applicable to other systems. Simulations of this type are computationally demanding at present,[32] but further developments in software and hardware will undoubtedly increase their efficiency. The ability to compute ΔG_{bind} values accurately will permit not only the optimization of lead compounds but also the design and evaluation of completely novel protein ligands.

Acknowledgment. We thank Tjerk Straatsma for providing the ARGOS program and Valère Lounnas and Razif Gabdoulline for critical reading of the manuscript. This work was supported in part by the EU Biotechnology Program (BIO2-CT94-2060).

JA9738539

(29) Bash, P. A.; Ho, L. L.; MacKerell, A. D.; Levine, D.; Hallst, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698−3703.

(30) Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins* **1991**, *11*, 281−296.

(31) Straatsma, T. P.; Berendsen, H. J. C.; Stam, A. J. *Mol. Phys.* **1986**, *57*, 89−95.

(32) The molecular dynamics simulations for this work required the equivalent of 48 days computation on a single R8000 processor of a Silicon Graphics Power Challenge Computer.