J|A|C|S
ARTICLES
Published on Web 03/12/2004

Proton Binding to Proteins: pK_a Calculations with Explicit and Implicit Solvent Models

Thomas Simonson,*,† Jens Carlsson,‡,§ and David A. Case*,‡

Contribution from the Laboratoire de Biochimie (UMR7654 du CNRS), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France and Department of Molecular Biology, Scripps Research Institute, La Jolla, California

Received November 25, 2003; E-mail: thomas.simonson@polytechnique.fr; case@scripps.edu

Abstract: Ionizable residues play important roles in protein structure and activity, and proton binding is a valuable reporter of electrostatic interactions in these systems. We use molecular dynamics free energy simulations (MDFE) to compute proton pK_a shifts, relative to a model compound in solution, for three aspartate side chains in two proteins. Simulations with explicit solvent and with an implicit, dielectric continuum solvent are reported. The implicit solvent simulations use the generalized Born (GB) model, which provides an approximate, analytical solution to Poisson's equation. With explicit solvent, the direction of the pK_a shifts is correct in all three cases with one force field (AMBER) and in two out of three cases with another (CHARMM). For two aspartates, the dielectric response to ionization is found to be linear, even though the separate protein and solvent responses can be nonlinear. For thioredoxin Asp26, nonlinearity arises from the presence of two substates that correspond to the two possible orientations of the protonated carboxylate. For this side chain, which is partly buried and has a large pK_a upshift, very long simulations are needed to correctly sample several slow degrees of freedom that reorganize in response to the ionization. Thus, nearby Lys57 rotates to form a salt bridge and becomes buried, while three waters intercalate along the opposite edge of Asp26. Such strong and anisotropic reorganization is very difficult to predict with Poisson−Boltzmann methods that only consider electrostatic interactions and employ a single protein structure. In contrast, MDFE with a GB dielectric continuum solvent, used for the first time for pK_a calculations, can describe protein reorganization accurately and gives encouraging agreement with experiment and with the explicit solvent simulations.

1. Introduction

The pH-dependent properties of proteins are of great importance for their structure and function.[1−4] Ionizable residues play key roles in such fundamental processes as protein folding, substrate binding, enzyme reactions, and redox behavior. Also, since proton binding is strongly influenced by electrostatic interactions with the protein and the surrounding solvent, pK_a measurements provide a valuable probe of protein electrostatics. Among the first applications of protein engineering were measurements of pK_a shifts and redox potential shifts associated with point mutations.[5,6] It was found that the electrostatic coupling between protein side chains depends not only on their distance, but also on their location in the protein and their proximity to solvent. Theoretical calculations, performed before protein engineering was possible, had already shown that such behavior was consistent with a simple continuum dielectric model.[7] The effect of protein−ligand and protein−protein binding on titration reactions has also been studied. The assembly of certain viruses is a striking example, where the ambient pH controls the overall process, because protons are bound by pairs of side chains belonging to adjacent coat protein subunits.[8] Conformational changes, such as partial or complete protein unfolding, can also lead to proton binding or release, providing the basis for the pH-dependence of protein stability.[3]

The effects mentioned above arise because proton binding depends sensitively on the environment of the titrating group. Three main, competing contributions are at work: the partial desolvation of the titrating residue in the protein environment, its interactions with charged and polar protein groups, and the ability of these groups to reorganize when the protonation state changes. Thus, ionizable residues buried in the interior of

† Laboratoire de Biochimie (UMR7654 du CNRS), Department of Biology, Ecole Polytechnique.

‡ Department of Molecular Biology, Scripps Research Institute.

§ Current address: Department of Molecular Biology, University of Uppsala, Sweden..

(1) Perutz, M. Electrostatic effects in proteins. *Science* **1978**, *201*, 1187−1191.
(2) Warshel, A.; Russell, S. Calculations of electrostatic effects in biological systems and in solutions. *Q. Rev. Biophys.* **1984**, *17*, 283−342.
(3) Schaefer, M.; Vlijmen, H. v.; Karplus, M. Electrostatic contributions to molecular free energies in solution. *Adv. Prot. Chem.* **1998**, *51*, 1−57.
(4) Simonson, T. Electrostatics and dynamics of proteins. *Rep. Prog. Phys.* **2003**, *66*, 737−787.
(5) Sternberg, M.; Hayes, F.; Russell, A.; Thomas, P.; Fersht, A. Prediction of electrostatic effects of engineering of protein charges. *Nature* **1987**, *330*, 86−88.
(6) Varadarajan, R.; Zewert, T.; Gray, H.; Boxer, S. Effects of buried ionizable amino acids on the reduction potential of recombinant hemoglobin. *Science* **1989**, *243*, 69−72.
(7) Hill, T. On intermolecular and intramolecular interactions between independent pairs of binding sites in proteins and other molecules. *J. Am. Chem. Soc.* **1956**, *78*, 5529.
(8) Caspar, D.; Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quantum Biol.* **1962**, *27*, 1−24.

10.1021/ja039788m CCC: $27.50

proteins can have substantial pK_a shifts compared to the isolated amino acid in solution. This is due to the less polar environment in most of the protein core, which, except for specific regions,[9] does not provide strong favorable interactions with the titrating amino acid in its ionized state. In a continuum dielectric picture, the protein core would be viewed as a low-dielectric medium, which is unfavorable for introducing a net charge.[7,10,11] On the other hand, polar or positively charged protein groups nearby could favor the unprotonated state of a carboxylate group, for example, and hence downshift its pK_a value. Such polar or charged groups are common in functional regions such as enzyme active sites.[9,12,13]

A more subtle effect is that of protein reorganization when the protonation state changes. For ionizable side chains that extend into solvent, this effect is small compared to reorganization of the solvent. But in functionally interesting regions, there are often ionizable side chains that are partly buried,[12,13] so that the protein response to their ionization is important. This reorganization is a microdielectric effect,[2,14,15] which can take the form of small reorientations of polar groups, similar to what would be seen in a simple dielectric medium. In other cases, ionization can also lead to an abrupt conformational switching, such as the making or breaking of a salt bridge.[16] An example is given by Asp26 in thioredoxin, below.

Although the above effects are essential, one should keep in mind that in general, proton binding is also affected by nonelectrostatic interactions. For example, the amount of water nearby, the extent of burial, and the flexibility of polar side chains around a titrating residue all depend on a balance of interactions, including van der Waals and hydrophobic interactions. Examples are seen in the systems studied below.

In this work, we use theoretical methods to compute the contributions of solvent, protein, and their reorganization to three carboxylate pK_a's in two proteins. Asp26 in thioredoxin, with a pK_a of 7.5,[17] has one of the largest upward pK_a shifts among protein carboxylates.[18] Asp20 in thioredoxin has an unshifted pK_a of 4.[18] Asp14 in ribonuclease A, with a pK_a of 2, has one of the lower carboxylate pK_a's.[18,19] Each of these residues titrates approximately independently of surrounding ionizable groups, in the sense that the ionization state of the other groups can be assumed fixed while the residue of interest binds a proton at a pH close to its pK_a. This simplifies the computational problem considerably. The possibility of multiple sites titrating in a concerted way, while important for other problems,[20] will not be considered here.

Several theoretical approaches are available for pK_a prediction. One popular approach is based on solving the linearized Poisson−Boltzmann equation (PB), using a purely electrostatic, continuum dielectric model for both protein and solvent.[10,21−24] These calculations are often based on just one conformation of the protein; e.g., the Xray structure. They do not explicitly account for the structural changes of the protein when the titrating proton binds. Rather, the structural relaxation of the protein (and the solvent) is modeled implicitly, as a rearrangement of polarization charge, governed by the protein and solvent dielectric constants.[4,25] This model works well for simple cases: e.g., solvent-exposed titrating groups with little or no pK_a shift; titrating groups in a rigid environment allowing little or no protein reorganization. But it has been shown to fail in several more complicated and interesting cases.[26,27] Some of the difficulties involved have been discussed in detail in the recent literature.[15,27−30]

Because a pK_a is the free energy to deprotonate a titratable site, the most fundamental way to calculate it is by molecular dynamics free energy simulations (MDFE).[31−33] Indeed, the first application of free energy simulations to a protein was a pK_a calculation.[31] MD simulations provide a detailed description, both of the protein and solvent structure and of their relaxation in response to the titrating charge, including electrostatic and nonelectrostatic effects. However, few studies have employed MDFE. Rather, the emphasis in the literature has been on simplified, PB or Langevin dipole models.[34,35] An intermediate

(9) Warshel, A. What about protein polarity? *Nature* **1987**, *330*, 15−17.
(10) Tanford, C.; Kirkwood, J. Theory of protein titration curves. General equations for impenetrable spheres. *J. Am. Chem. Soc.* **1957**, *79*, 5333−5339.
(11) Warshel, A. Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5250.
(12) Elcock, A. Prediction of functionally important residues based solely on the computed energetics of protein structure. *J. Mol. Biol.* **2001**, *312*, 885−896.
(13) Ondrechen, M.; Clifton, J.; Ringe, D. THEMATICS: a simple computational predictor of enzyme function from structure. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12 473−12 478.
(14) Simonson, T.; Perahia, D. Microscopic dielectric properties of cytochrome *c* from molecular dynamics simulations in aqueous solution. *J. Am. Chem. Soc.* **1995**, *117*, 7987−8000.
(15) Schutz, C.; Warshel, A. What are the dielectric 'constants' of proteins and how to validate electrostatic models? *Proteins* **2001**, *8*, 211−217.
(16) Eberini, I.; Baptista, A.; Gianazza, E.; Fraternali, F.; Beringhelli, T. Reorganization in apo- and holo-β-lactoglobulin upon protonation of Glu89: molecular dynamics and pK_a calculations. *Proteins* **2004**, *54*, 744−758.
(17) Langsetmo, K.; Fuchs, J.; Woodward, C. The conserved, buried aspartic acid in oxidized *Escherichia coli* thioredoxin has a pK_a of 7.5. Its titration produces a related shift in global stability. *Biochemistry* **1991**, *30*, 7603−7609.
(18) Forsyth, W.; Antosiewicz, J.; Robertson, A. Empirical relationships between protein structure and carboxyl pK_a values in proteins. *Proteins* **2002**, *48*, 388−403.
(19) Baker, W.; Kintanar, A. Characterization of the pH titration shifts of ribonuclease A by one and two-dimensional nuclear magnetic resonance spectroscopy. *Arch. Biochem. Biophys.* **1996**, *327*, 189−199.
(20) Onufriev, A.; Case, D.; Ullmann, M. A novel view of pH titration in biomolecules. *Biochemistry* **2001**, *40*, 3413−3419.
(21) Bashford, D.; Karplus, M. The pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **1990**, *29*, 10 219−10 225.
(22) Antosiewicz, J.; McCammon, J.; Gilson, M. Prediction of pH dependent properties of proteins. *J. Mol. Biol.* **1994**, *238*, 415−436.
(23) Demchuk, E.; Wade, R. Improving the continuum dielectric approach to calculating pK_a's of ionizable groups in proteins. *J. Phys. Chem.* **1996**, *100*, 17 373−17 387.
(24) Havranek, J.; Harbury, P. Tanford-Kirkwood electrostatics for protein modeling. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11 145−11 150.
(25) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1−20.
(26) Langsetmo, K.; Fuchs, J.; Woodward, C.; Sharp, K. Linkage of thioredoxin stability to titration of ionizable groups with perturbed pK_a. *Biochemistry* **1991**, *30*, 7609−7614.
(27) Dillet, V.; Dyson, J.; Bashford, D. Calculations of electrostatic interactions and pK_a's in the active site of *Escherichia coli* thioredoxin. *Biochemistry* **1998**, *37*, 10 298−10 306.
(28) Sham, Y.; Chu, Z.; Warshel, A. Consistent calculations of pK_a's of ionizable residues in proteins: semi-microscopic and microscopic approaches. *J. Phys. Chem. B* **1997**, *101*, 4458−4472.
(29) Simonson, T.; Archontis, G.; Karplus, M. A Poisson−Boltzmann study of charge insertion in an enzyme active site: the effect of dielectric relaxation. *J. Phys. Chem. B* **1999**, *103*, 6142−6156.
(30) Archontis, G.; Simonson, T. Dielectric relaxation in an enzyme active site: molecular dynamics simulations intepreted with a macroscopic continuum model. *J. Am. Chem. Soc.* **2001**, *123*, 11 047−11 056.
(31) Warshel, A.; Sussman, F.; King, G. Free energy changes in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry* **1986**, *25*, 8368−8372.
(32) Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395.
(33) Simonson, T. Free energy calculations. In *Computational Biochemistry & Biophysics*; O. Becker, A. Mackerell, Jr.; B. Roux; M. Watanabe, Eds. Marcel Dekker: NY, 2001, ch. 9.
(34) Warshel, A.; Papazyan, A. Electrostatic effects in macromolecules: fundamental concepts and practical modelling. *Curr. Op. Struct. Biol.* **1998**, *8*, 211−217.

approach employed MD with explicit solvent to sample conformations of the ionized and neutral states, along with a simplified model to evaluate their energetics.[28] The present study therefore remains one of the few attempts to calculate pK_a shifts in proteins with MD free energy simulation methods. Compared to the pioneering early studies of Warshel and co-workers,[31,36] it benefits from more recent force fields, orders of magnitude more computer power (which turns out to be important) and, consequently, larger simulation models, along with Particle Mesh Ewald electrostatic treatments.[37,38] We focus on three aspartic acids, two of which are especially difficult cases, since they have highly perturbed pK_a's compared to the free amino acid in solution. Good agreement is found overall for the existence and direction of the pK_a shifts, although the quantitative agreement is not perfect.

More sophisticated PB methods also exist, which include protein relaxation explicitly;[39−48] see refs 4 and 25 for recent reviews. One such method is used here, and compared to the explicit solvent calculations. Unlike the 'standard' PB continuum approaches discussed above, we take an approach where the protein flexibility and dielectric relaxation are treated explicitly, by molecular dynamics simulation, whereas the solvent is modeled as a dielectric continuum. The method is made efficient by using the generalized Born (GB) model,[25,46,47] where the numerical solution of the Poisson equation is replaced by an approximate analytical solution. While most continuum dielectric pK_a calculations have at least one adjustable parameter (the protein dielectric constant), this approach does not. It is 1−2 orders of magnitude more efficient than MDFE with explicit solvent. This work and another recent study[49] represent the first application of the GB model to protein pK_a calculations. The results turn out to be qualitatively correct, in very good agreement with the explicit solvent results, although again, the detailed agreement with experiment is not perfect.

In addition to the main issues described above, our calculations allow us to address several secondary, technical questions that are nevertheless of significant interest. We directly compare simulations with two widely used protein force fields (AMBER and CHARMM), and we compare two variants of the GB model.[50−52]

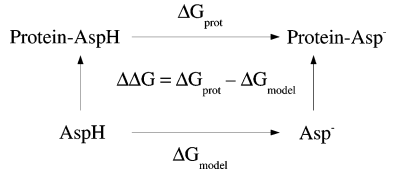

Figure 1. Thermodynamic cycle to analyze pK_a shifts. The upper leg represents proton binding to an Asp side chain in a protein. The lower leg represents proton binding to a model compound in solution (aspartate with neutral blocking groups; Figure 2). The double free energy difference between the two legs is $\Delta\Delta G = \Delta G_{prot} - \Delta G_{model}$.

2. Methods

2.1 Calculating pK_a Shifts from Free Energies. The method used here was first proposed by Warshel and co-workers.[31] It is suitable for calculating pK_a shifts; i.e., the pK_a difference between an ionizable group in a protein (such as an Asp side chain) and the same group in solution (aspartate). We therefore consider the two proton binding reactions shown in Figure 1. The equilibrium constant K_a for either reaction is related to the standard reaction free energy ΔG by

$$\Delta G = -kT \log K_a \tag{1}$$

Therefore

$$pK_a = -\log_{10} K_a = \frac{1}{2.303kT}\Delta G \tag{2}$$

and

$$pK_{a,prot} = pK_{a,model} + \frac{1}{2.303kT}\Delta\Delta G, \tag{3}$$

where p$K_{a,prot}$ and p$K_{a,model}$ are the pK_a's, respectively, of the titrating group in the protein and the reference molecule in solution, and $\Delta\Delta G = \Delta G_{prot} - \Delta G_{model}$ is the double free energy difference for the thermodynamic cycle in Figure 1.

2.2 Classical Mechanical Model of Proton Binding. We adopt a simplified description of the proton binding reaction,[31] consistent with the molecular mechanics model employed here (and in most biomolecular simulation studies). The proton, like the other atoms in the system, is treated as a classical mechanical particle, bearing a partial charge, and interacting with the other atoms through Coulombic terms, stereochemical terms, and possibly van der Waals terms (although the force fields used here do not actually assign van der Waals interactions to the aspartate side chain proton). Proton binding is then described as the introduction of a new particle into the molecular mechanics model. Most pK_a calculations based on a simple continuum dielectric model have modeled the proton binding by a rearrangement of partial charges on the 'heavy' (non-hydrogen) atoms of the titrating side chains.[3,21]

(35) Mehler, E.; Guarneri, F. A self-consistent, microenvironment modulated screened Coulomb potential approximation to calculate pH-dependent effects in proteins. *Biophys. J.* **1999**, *77*, 3−22.
(36) Warshel, A. *Computer modelling of chemical reactions in enzymes and solutions*: John Wiley: New York, 1991.
(37) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10 089−10 092.
(38) Sagui, C.; Darden, T. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155−179.
(39) Georgescu, E.; Alexov, E.; Gunner, M. Combining conformational flexibility and continuum electrostatics for calculating pk$_a$'s in proteins. *Biophys. J.* **2002**, *83*, 1731−1748.
(40) You, T.; Bashford, D. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. *Biophys. J.* **1995**, *69*, 1721−1733.
(41) Beroza, P.; Case, D. Including side chain flexibility in continuum electrostatic calculations of protein titration. *J. Phys. Chem.* **1996**, *100*, 20 156−20 163.
(42) Alexov, E.; Gunner, M. Calculated protein and proton motions coupled to electron transfer: electron transfer from Q_A^- to Q_B in bacterial photosynthetic reaction centers. *Biochemistry* **1999**, *38*, 8253−8270.
(43) Gilson, M.; Davis, M.; Luty, B.; McCammon, J. Computation of electrostatic forces on solvated molecules using the Poisson−Boltzmann equation. *J. Phys. Chem.* **1993**, *97*, 3591−3600.
(44) Im, W.; Beglov, D.; Roux, B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson−Boltzmann equation. *Comput. Phys. Comm.* **1998**, *111*, 59−75.
(45) M. Marchi and D. Borgis and N. Levy and P. Ballone. A dielectric continuum molecular dynamics method. *J. Chem. Phys.* **2001**, *114*, 4327−4385.
(46) Still, W. C.; Tempczyk, A.; Hawley, R.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(47) Bashford, D.; Case, D. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(48) Hassan, S.; Zhang, D.; Mehler, E.; Weinstein, H. A fast and general continuum approach for describing electrostatic effects in molecular dynamics simulations of biomolecules. *Biophys. J.* **2002**, *82*, 2363.
(49) Lee, M.; Salsbury Jr.; F.; Brooks, III, C. Constant pH molecular dynamics using continuous titration coordinates. *Proteins*, in press, 2004.
(50) Onufriev, A.; Bashford, D.; Case, D. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712−3720.
(51) Schaefer, M.; Karplus, M. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(52) Schaefer, M.; Bartels, C.; Karplus, M. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **1998**, *284*, 835−847.

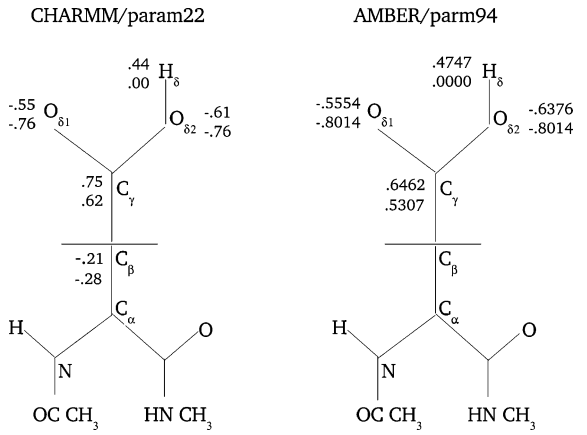

Figure 2. Schematic view of the model compound (2*N*-acetyl-1*N*-methyl-aspartic acid-1-amide), showing the partial charges in the protonated state (above) and the ionized state (below). Left panel: CHARMM values; right panel: AMBER values (only shown for atoms that change between the two states).

The model used here is more general, and should be more accurate, since the proton is explicitly represented.

The two force fields employed (AMBER/parm94[53] and CHARMM22[54]) use fixed partial charges; ie, electronic polarizability is treated in a simple, mean field way. Explicit electronic polarizability could be included, either at a classical mechanical level, through a polarizable force field,[55] or quantum mechanically, through a so-called QM/MM model[56] or a valence bond model.[36] It would also be straightforward to treat the nuclear degrees of freedom of the proton quantum mechanically, using the Feynmann path integral representation, adapted for molecular simulations by Chandler and Wolynes.[57] However, for pK_a shifts, the simple, classical mechanical description used here (and in most previous work) is a reasonable compromise between accuracy and efficiency.

As a model compound (lower leg of the cycle in Figure 1), we used aspartic acid with *N*-acetyl and *N*-methylamide blocking groups (Figure 2). Its pK_a value is equal to 4.0.[58] The charge distributions in the protonated and ionized states are shown in Figure 2, for both the AMBER and CHARMM force fields. In both force fields, the change of charge is localized on the carboxylate group, along with the CB atom in the CHARMM case.

2.3 Molecular Dynamics Free Energy Simulations. In a free energy simulation, one gradually transforms the system of interest from an initial state A (protonated Asp) to a final state B (ionized Asp), by modifying the energy function U.[33] The simplest scheme is to make U a linear function of a progress variable, or 'coupling parameter' λ

$$U(\lambda) = (1 - \lambda)U_A + \lambda U_B \qquad (4)$$

where U_A and U_B are the energy functions corresponding to the initial and final states. Varying λ from zero to one takes us from A to B; the intermediate values $0 < \lambda < 1$ correspond to a hybrid system that is a mixture of A and B. In both the NVT and NPT ensembles, the free energy derivative with respect to λ has the form

$$\partial G/\partial\lambda\ (\lambda) = \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda = \langle U_B - U_A \rangle_\lambda = \sum_i \delta q_i \langle V_i \rangle_\lambda \qquad (5)$$

where the brackets represent an ensemble average in the ensemble corresponding to $U(\lambda)$; δq_i is the charge increment of atom i going from the reactant to the product state; V_i is the electrostatic potential on atom i, and the sum on the right is over the atoms that carry the new charge (the Asp carboxylate group; see Figure 2). Borrowing the language of electron transfer theory, the quantity that is averaged, $U_B - U_A$, is referred to as the energy gap.[59] It represents the energy to perform a virtual ionization reaction with the system frozen in its instantaneous conformation.

Performing simulations for a series of values of λ between zero and one, one can then obtain the free energy change ΔG by numerical integration of $\partial G/\partial\lambda$. This procedure is referred to as 'thermodynamic integration' (TI). In the CHARMM free energy simulations, it is actually the carboxylate charges that are transformed linearly, instead of the energy function itself. This leads to an additional, quadratic term in the free energy derivative (not shown in eq 5). This term is essentially constant and identical for the protein and the model compound (see ref 4 for details), and can therefore be ignored here. Ewald 'self' terms,[60] due to coupling between the charge increments in neighboring periodic cells, can be ignored for the same reason.[61]

A different numerical route to the free energy change ΔG is to sum the free energy increments ΔG_i between successive values of λ

$$\Delta G = \sum_{i=0}^{n-1} \Delta G_i = -kT \sum_{i=0}^{n-1} \log\left\langle \exp\left(-\frac{U(\lambda_{i+1}) - U(\lambda_i)}{kT} \right) \right\rangle_i \qquad (6)$$

where the first value of λ is $\lambda_0 = 0$; the last is $\lambda_n = 1$, and the brackets represent an ensemble average with the energy function $U(\lambda_i)$. This procedure is referred to as 'thermodynamic perturbation' (TP).

2.4 Starting Structures. The first reaction studied was the protonation of Asp26 of the oxidized form of thioredoxin from *Escherichia coli*, which has an experimental pK_a of 7.5 and titrates approximately independently of other ionizable groups.[27] To calculate the standard reaction free energy, the free energy simulation is performed in conditions that mimic a pH of 7.5. The initial structure was a 1.68 Å resolution X-ray structure (PDB code 2TRX) determined at pH = 4.1.[62] The unique histidine (His6) was doubly protonated. This residue is solvent-exposed, 10 Å away from Asp26, so that its protonation state is not expected to affect Asp26 strongly. All other titratable residues besides Asp26 were assigned their most common ionization state at pH 7.5. Except for Asp26, the ionization state of all residues was assumed constant throughout the free energy simulations.

The second ionizable residue studied was Asp20 of thioredoxin. Its pK_a is not known exactly, but can be inferred from Asp20 in the highly homologous human thioredoxin, which has a 'normal' pK_a of 4.0.[18] In this case, the free energy simulation was performed with Asp26 protonated and His6 doubly protonated. This should be representative of the protein's state at pH 4. Since all other carboxylate groups (besides Asp26) are rather distant from Asp20, their exact ionization state should not affect the Asp20 proton binding. They were assumed ionized

(53) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(54) Mackerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W.; Roux, B.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(55) Lamoureux, G.; MacKerell Jr.; A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119*, 5185−5197.
(56) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227.
(57) Chandler, D. Quantum theory of solvation. *J. Phys. Chem.* **1984**, *88*, 3400−3407.
(58) Kyte, J. *Structure in Protein Chemistry*; Garland Publishing: New York, 1995.
(59) Warshel, A. Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron transfer and proton-transfer studies. *J. Phys. Chem.* **1982**, *86*, 2218−2224.
(60) Hummer, G.; Pratt, L.; Garcia, A. Free energy of ionic hydration. *J. Phys. Chem.* **1996**, *100*, 1206−1215.
(61) Simonson, T. Gaussian fluctuations and linear response in an electron-transfer protein. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6544−6549.

throughout the CHARMM simulations and protonated throughout the AMBER simulations.

The third reaction studied was the protonation of Asp14 of human pancreatic ribonuclease A (RNase A), whose experimental pK_a is 2.0. The initial structure was the 1.05 Å X-ray structure, determined at pH 5.2 (PDB code 1KF2). This structure is very similar to structures determined at a near-neutral pH (PDB codes 1KF4, 1KF5). The four histidines were doubly protonated. Despite the low pK_a, most of the simulations were done with all carboxylate groups other than Asp14 in their ionized state. Since they are all solvent exposed and rather far from Asp14, this should not have a large effect on the calculated Asp14 pK_a. GB simulations were performed with them either protonated or deprotonated (see below); the results are indeed similar.

2.5 Explicit Solvent Simulation Setup. Two sets of free energy simulations were performed with explicit solvent. First, all three pK_a's were calculated with the AMBER program,[63] using the AMBER/parm94 force field[53] and the TIP3P water model.[64] Periodic boundary conditions were used, with an octahedral unit cell. The cell volume was determined by 160−300 ps equilibration runs under NPT conditions, with a temperature of 298 K. After that, simulations were continued in the NVE ensemble. Long-range electrostatic interactions were treated by the Particle Mesh Ewald method (PME).[38] Covalent bonds to hydrogens were constrained with SHAKE;[65] the time step was 1 fs. For the thioredoxin runs, the octahedral cell had a volume of about 181 000 Å^3, and contained 4757 waters and four Na^+ ions (Asp26 runs) or 4753 waters and three Na^+ ions (Asp20 runs). For the ribonuclease A runs, the volume was 267 923 Å^3, with 7324 waters and nine Cl^- ions. Run lengths are indicated in the Results section.

All three pK_a's were also calculated with the CHARMM program,[66] using the CHARMM/parm22 force field.[54] The periodic cell was cubic. Temperature and pressure were maintained at 295 K and 1 atm by a Nosé-Hoover thermostat and barostat.[67,68] For thioredoxin, the unit cell edge was about 59 Å; the cell included 6379 waters and three Na^+ ions. For RNase A, the cell edge was about 67 Å; it included 9422 waters and nine Cl^- ions. The first 100−140 ps of dynamics included weak harmonic restraints toward the crystal structure, which were gradually removed. Other conditions were the same as for the AMBER runs. The model compound was simulated in the same conditions, with a box of the same size.

2.6 Generalized Born Continuum Solvent Model. The PME simulation models described above treat all of the solvent explicitly. This is the most accurate approach, but also the most expensive. A more approximate route is to treat all the solvent as a dielectric continuum. This is the basis of the Generalized Born (GB) approach,[25,46,47] summarized below.

The electrostatic interaction between two charges i, j includes both a direct, Coulomb term and a contribution from the solvent, polarized by the solute charges. Treating the solvent as a linear, homogeneous, dielectric medium, the total electrostatic energy has the form

$$E^{\mathrm{elec}} = \frac{1}{2}\sum_{i\neq j}\frac{q_i q_j}{r_{ij}} + \frac{1}{2}\sum_{ij} g_{ij} \tag{7}$$

where the sums are over all pairs of protein charges. The term g_{ij} in the second sum is a Green's function, representing the interaction between a protein charge q_i and the solvent polarization induced by another charge, q_j. In the generalized Born model,[46] this term is approximated by

$$g_{ij} = g(r_i, r_j) = \frac{\tau q_i q_j}{(r_{ij}^2 + b_i b_j \exp[-r_{ij}^2/4b_i b_j])^{1/2}} \tag{8}$$

where $r_{ij} = |r_i - r_j|$, $\tau = 1/\epsilon_w - 1$, ϵ_w is the solvent dielectric constant (80 at room temperature), and b_i, b_j are effective 'solvation radii' of the charges i, j. The interaction term g_{ij} depends explicitly on the atomic positions r_i, r_j, and implicitly on all the other atomic positions, through the solvation radii. Indeed, the solvation radius b_i is determined by the 'self' energy E_i^{self} of charge i

$$E_i^{\mathrm{self}} = \frac{1}{2} g_{ii} = \frac{\tau q_i^2}{2b_i} \tag{9}$$

E_i^{self} is the interaction energy between q_i and the polarization it creates in the solvent. In practice, b_i is roughly equal to the shortest distance between q_i and the protein surface. In the GB model, it is approximated by a simple, analytical function of the positions of all the solute atoms (including those that have a zero partial charge): $b_i = b_i(r_1, r_2, ..., r_N)$. Different GB variants use different functional forms. In most variants, including the ones considered here, the self-energy takes the form of a pairwise sum over atoms

$$E_i^{\mathrm{self}} = \sum_j E_{ij}^{\mathrm{self}}(r_i, r_j) \tag{10}$$

2.7 Free Energy Simulations with the GB Model. We have done calculations using two variants of the GB model. The GB/ACE variant was developed by Schaefer and Karplus,[51] and first used for protein MD simulations by Calimet et al.[69] The GB/OBC variant was developed by Onufriev, Bashford and Case,[50,70] starting from a self-energy model by Ponder, Truhlar, and co-workers.[71,72] These variants contain essentially the same physics, but use somewhat different numerical techniques. The details of the self-energy approximation with each variant are given in the original papers.[51,71]

Free energy simulations were performed to obtain the pK_a estimates. The appropriate charges were made to vary with the coupling parameter λ, and the standard free energy change ΔG was obtained by TI or TP, as above. The simulations were started from the crystal structure. Harmonic restraints were used during equilibration and gradually released. The temperature was controlled by a weak velocity coupling[73] in the GB/OBC runs (done with AMBER), or by a Langevin dynamics method in the GB/ACE runs (done with CHARMM). The λ values employed were 0, 0.5 and 1 with CHARMM, and 0, 0.1127, 0.5, 0.88729, and 1 with AMBER. With CHARMM, 5 ns were run for each λ value; with AMBER, 3 ns were run, giving the same total amount of MD. The GB/ACE simulations (with CHARMM) employed the param19 force field,[66] which treats aliphatic hydrogens implicitly,

(62) Katti, S.; Lemaster, D.; Eklund, H. Crystal structure of thioredoxin from *Escherichia coli* at 1.68 å resolution. *J. Mol. Biol.* **1990**, *212*, 167.
(63) Case, D.; Pearlman, D.; Caldwell, J., III, T. C.; Ross, W.; Simmerling, C.; Darden, T.; Merz, K.; Stanton, R.; Cheng, A.; Vincent, J.; Crowley, M.; Tsui, V.; Radmer, R.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G.; Singh, U.; Weiner, P.; Kollman, P. *AMBER 6*; University of California, San Francisco, 1999.
(64) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(65) Ryckaert, J.; Ciccotti, G.; Berendsen, H. Numerical integration of the Cartesian equations of motion for a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(66) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(67) Nose, S. A unified formulation of the constant temperature molecular dynamics method. *J. Chem. Phys.* **1984**, *81*, 511−519.
(68) Hoover, W. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(69) Calimet, N.; Schaefer, M.; Simonson, T. Protein molecular dynamics with the Generalized Born/ACE solvent model. *Proteins* **2001**, *45*, 144−158.
(70) Onufriev, A.; Bashford, D.; Case, D. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins* 2004, in press.
(71) Hawkins, G.; Cramer, C.; Truhlar, D. Pairwise descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(72) Ponder, J. *Tinker*; Washington University School of Medicine, 2000.
(73) Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *811*, 3684−3690.

Table 1. Thioredoxin Asp26: Explicit Solvent MDFE Results[a]

	CHARMM			AMBER	
	model	Asp26[b]	Asp26[c]	model	Asp26
run length (ns)	6	9.5	8	3	18
$\partial G/\partial\lambda$ ($\lambda = 0$)	19.3(1.4)	19.3(6.8)	16.6(4.6)		
$\partial G/\partial\lambda$ ($\lambda = 0.11270$)				−1.3(0.8)	−3.1(12.0)
$\partial G/\partial\lambda$ ($\lambda = 0.5$)	−58.4(0.8)	−42.0(0.4)	−44.8(2.0)	−75.3(0.2)	−64.5(4.6)
$\partial G/\partial\lambda$ ($\lambda = 0.88279$)				−148.6(2.0)	−131.4(7.8)
$\partial G/\partial\lambda$ ($\lambda = 1$)	−144.6(1.4)	−122.0(4.8)	−124.2(2.2)		
ΔG	−60.5(0.6)	−46.7(2.0)	−49.3(1.6)	−75.1(1.1)	−66.0(3.9)
$\Delta\Delta G$	0	13.8(2.0)	10.9(1.8)	0	9.1(4.1)
$\Delta\Delta G$ (exper.)		4.8	4.8		4.8

[a] All energies in kcal/mol. Statistical uncertainty in parentheses. For the derivatives, it is estimated as twice the standard deviation of block averages, each trajectory being divided into four blocks. For the free energies, calculated by trapezoidal (CHARMM) or Gaussian (AMBER) integration, simple propagation of error is assumed. [b] Forward run (starting from the protonated state, $\lambda = 0$). [c] Backward run (starting from the ionized state, λ=1).

instead of the all-hyrogen param22 force field used for the explicit solvent runs. The GB/ACE parametrization was optimized previously[69] for use with param19. The GB/OBC simulations (with AMBER) employed a slightly modified version of the AMBER/parm99 force field, optimized prevously for GB simulations.[74] No parameter optimization (or fitting) was done in this work.

3. Results

The most complex and interesting ionization reaction is that of Asp26 in thioredoxin, which we describe first, focusing on the explicit solvent simulations. The ionization of Asp20 in thioredoxin and Asp14 in RNase A are then described more briefly. Finally, we describe the implicit solvent results for all three ionization reactions.

3.1 Thioredoxin Asp26. Explicit Solvent Simulations. Ionization Free Energy. The Asp26 ionization reaction was simulated in three sets of free energy simulations (three 'runs'). A run with the AMBER force field and Ewald summation included 18 ns of dynamics. A forward and a backward run were done with the CHARMM force field and Ewald summation, totalling 17 ns of dynamics. These simulation lengths were necessary to correctly sample the important degrees of freedom, as described below. Ionization of the model compound (Asp with neutral blocking groups; Figure 2) was simulated with AMBER over 3 ns, and with CHARMM over 6 ns. The CHARMM run for the model compound was fully converged after 3 ns; ie, doubling the simulation time for each λ value did not change the result. The derivative of the free energy, $\partial G/\partial\lambda$, is listed vs λ in Table 1 for the model compound and the protein. A representative time series for the energy gap, $U_B - U_A$ (eq 5), is shown in Figure 3, illustrating the convergence of $\partial G/\partial\lambda$. The free energy change is obtained by numerical integration of the derivative.

From the different runs, the estimated free energy change $\Delta\Delta G$ relative to the model compound is 9.1 kcal/mol (AMBER), 10.9 kcal/mol (CHARMM, backward run) or 13.8 kcal/mol (CHARMM, forward run). The experimental result is 4.8 kcal/mol. Thus, while the simulations consistently predict the correct sign for the Asp26 pK_a shift, the magnitude is over twice the experimental one. The 6.6 kcal/mol difference from experiment is about three times the standard error, 2.3 kcal/mol (estimated from the scatter of the three runs). Therefore, it must arise in part from systematic errors in the force fields employed. For example, both force fields treat electronic polarizability implicitly, through a simple, mean field model.[53,54]

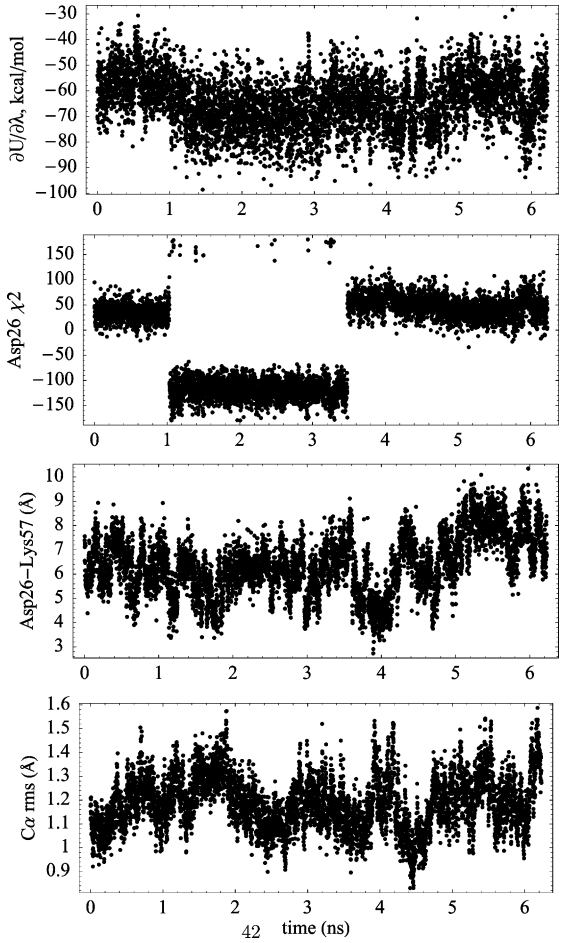

Figure 3. Representative data from the AMBER simulation of the half-ionized Asp26 state of thioredoxin. From top to bottom: time series of the energy gap $\partial U/\partial\lambda$, of the Asp26 χ_2 dihedral, of the Asp26(OD)-Lys57(NZ) distance (taking the OD oxygen that gives the shortest distance), and of the mean C_α rms deviation from the crystal structure.

Despite the length of the simulations, the scatter among the free energy runs is still significant. While the AMBER/CHARMM differences could in principle be dominated by systematic force field effects, they are of the same magnitude as the apparent statistical uncertainty of each run (in parentheses in Table 1). This suggests that the differences between runs are mainly due to conformational sampling differences. They arise because the deprotonation of Asp26 is a complex process,

(74) Simmerling, C.; Strockbine, B.; Roitberg, A. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **2002**, *124*, 11 258−11 259.

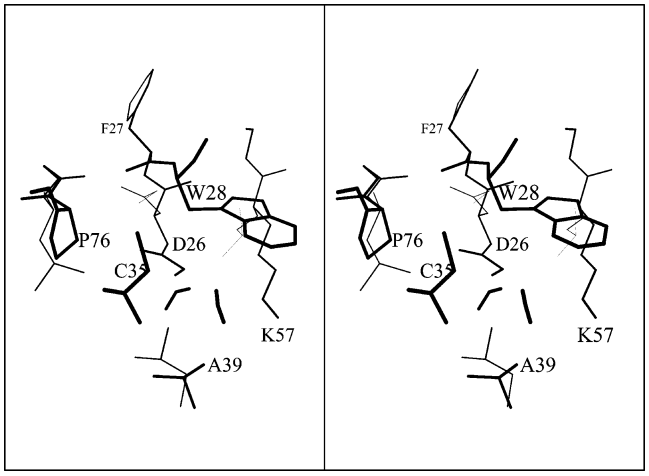

Figure 4. Environment of thioredoxin Asp26 in its neutral state (typical snapshot from the CHARMM simulation).

involving structural rearrangements in the protein and the solvent that take place on long, nanosecond, time scales. We describe these in the next two sections. The difference between the CHARMM forward and backward runs is greater than the statistical uncertainty of either run, and must be due to systematic 'hysteresis'; i.e., the forward run is too short to allow the ionized endpoint to become fully stabilized, thus yielding a higher pK_a.

Structural Reorganization: Carboxylate Orientation. Asp26 lies in a deep pocket on the protein surface, illustrated in Figure 4. Several slow degrees of freedom in the immediate vicinity of Asp26 are strongly coupled to the reaction coordinate λ (eq 4). Thus, the Asp26 χ_2 dihedral angle has a preferred orientation in the protonated state, but obviously not in the ionized state. In the CHARMM forward run, in the protonated state ($\lambda = 0$), Asp26 starts out with $\chi_2 \approx -120°$, with the proton pointing away from solvent, back into the pocket. After 800−900 ps of dynamics, the carboxylate flips by about 180°, so that the proton points outward (toward the right of Figure 4). This 'proton-outward' orientation is maintained for over 3 ns of simulation, suggesting that it is the favored orientation with the CHARMM force field at $\lambda = 0$. In the midpoint state ($\lambda = 0.5$), only the 'proton-outward' orientation is sampled.

In the AMBER simulations (which correspond to intermediate λ values: $\lambda = 0.11270$, 0.5, and 0.88729), the two Asp26 orientations are sampled about equally, with transitions every 1−2 ns (Figure 3). The free energy derivative or, equivalently, the electrostatic potential on Asp26, has a distinct dependency on the dihedral orientation, detailed in Table 2. For example, in the $\lambda = 0.5$ state, if we limit the averaging to the proton-outward (respectively, inward) orientation, the free energy derivative is estimated to be −62.0 kcal/mol (respectively, −68.4 kcal/mol). This dependency accounts for the large apparent statistical uncertainty of the AMBER run (4.1 kcal/mol; Table 1), compared to the CHARMM runs (∼2 kcal/mol). Since there are at least two χ_2 transitions in each simulation, we might assume the inward and outward populations are sampled at least qualitatively correctly, and simply average the data over the whole trajectory. However, from the data, it is actually possible to compute the statistical weights of each conformer in a rigorous way.[33,75,76] Indeed, treating the inward and outward conformers as two separate species in slow exchange, we can

Table 2. Thioredoxin Asp26 χ_2 Substate Free Energies[a]

	proton-inward	proton-outward	overall
$\partial G/\partial\lambda$ ($\lambda = 0.11270$)	5.0(0.5)	−10.7(3.4)	−3.1
$\partial G/\partial\lambda$ ($\lambda = 0.5$)	−68.4(1.7)	−62.0(4.0)	−64.5
$\partial G/\partial\lambda$ ($\lambda = 0.88279$)	−133.3(2.8)	−129.5(2.4)	−131.4
ΔG	−65.8(1.1)	−66.8(2.3)	−66.0(2.5)[b]

[a] AMBER run. All energies in kcal/mol. Statistical uncertainty in parentheses. For the derivatives, it is estimated as twice the standard deviation of block averages, each data set being divided into four blocks. [b] For the overall free energy, the uncertainty is estimated by propagation of error from the substate free energy uncertainties: $\sigma(\Delta G) = (\sigma(\Delta G_{in})^2 + \sigma(\Delta G_{out})^2)^{1/2}$. The substate free energies are actually combined through a nonlinear, Boltzmann average (eq 11); nevertheless, they have almost equal weights in the average so that the above 'linear' mixing of the errors is a good approximation.

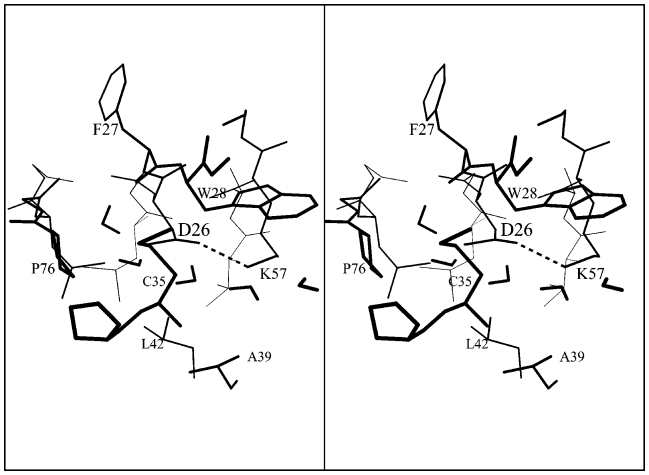

Figure 5. Environment of thioredoxin Asp26 in its ionized state (typical snapshot from the CHARMM simulation).

compute a free energy change ΔG for each one, giving $\Delta G_{in} = -65.9$ kcal/mol and $\Delta G_{out} = -66.9$ kcal/mol, respectively (Table 2). The product states for the two species have the same free energy, because for $\lambda = 1$, the two Asp26 orientations are equivalent. We may conclude that the inward orientation is favored by $\Delta G_{in} - \Delta G_{out} = 1$ kcal/mol in the reactant state. This contrasts with the CHARMM force field, which favors the outward state. Finally, we take the Boltzmann average of the inward and outward results

$$\Delta G = -kT \log[\exp(-\Delta G_{in}/kT) + \exp(-\Delta G_{out}/kT)] \quad (11)$$

to obtain $\Delta G = -66.4$ kcal/mol. This is very close to the result of the simple averaging, $\Delta G = -66.0$ kcal/mol (Table 1). This analysis shows that the real statistical uncertainty (2.5 kcal/mol; Table 2) is close to the uncertainty of the CHARMM runs.

Structural Reorganization: Salt Bridge Formation and Solvent Structure. A second important degree of freedom involves the positively charged Lys57 side chain. Lys57 does not interact closely with protonated Asp26 (Figure 4), but shifts several Ångstroms to make a salt bridge to ionized Asp26 (Figure 5). The salt bridge is almost completely absent up to, and including the midpoint of the reaction ($\lambda = 0.5$, 'half-ionized' Asp26; see third panel of Figure 3). It is present about half the time when $\lambda = 0.887\,29$ (AMBER run). It is present

(75) Tobias, D. J.; Brooks, C. L. Calculation of free energy surfaces using the methods of thermodynamic perturbation theory. *Chem. Phys. Lett.* **1987**, *142*, 472−476.
(76) Hummer, G.; Pratt, L.; Garcia, A. Multistate Gaussian model for electrostatic solvation free energies. *J. Am. Chem. Soc.* **1997**, *119*, 8523−8527.

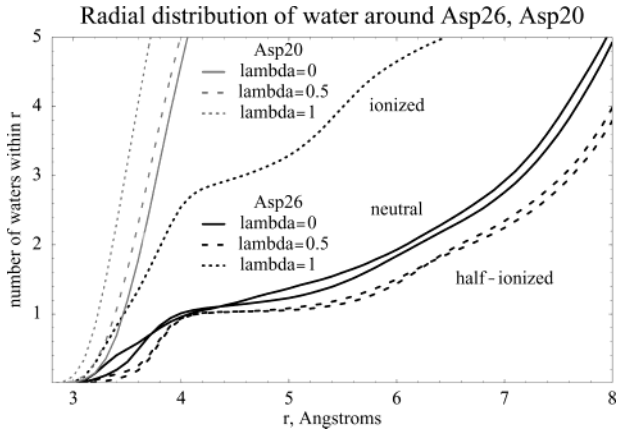

Figure 6. Radial distribution of water molecules around thioredoxin Asp26 and Asp20 in different states (neutral, half-ionized, ionized).

all the time when $\lambda = 1$ (fully ionized Asp26), as illustrated in Figure 5. The lack of a direct interaction in the protonated state was known from the 2TRX crystal structure.[62] However, the salt bridge found here in the product state could not be observed experimentally, because of the low occupancy of the product state at neutral pH.

A third important degree of freedom involves the solvent organization close to Asp26, which is completely different in the protonated and ionized states. Protonated Asp26 interacts directly with a single water molecule (Figure 4), which is the beginning of a chain of three waters in single file, linking Asp26 to bulk solvent. This organization agrees with the crystal structure.[62] The link is occasionally broken, as protein groups lining the chain come together and squeeze the waters out. This occurs during the early part of the $\lambda = 0$ window of the CHARMM run, for example. The groups involved are mainly the aliphatic part of Lys57, the Trp28 side chain above it, and the Ala39 methyl group below it; their opening/closing represents another slow, collective degree of freedom coupled to the reaction coordinate. To the left of Asp26 are Pro76 and the backbone of residues 76−79, none of which make a good hydrogen bond to the Asp26 side chain (regardless of the χ_2 orientation). Interestingly, in the half-ionized state ($\lambda = 0.5$), the solvent organization is very similar to that in the protonated state, as shown by the radial distribution functions plotted in Figure 6. As Asp26 goes from the protonated to the half-ionized state, its neighboring water molecule actually shifts away by about 0.3 Å and becomes noticeably more ordered.

Ionized Asp26, in contrast, is more extensively solvated. It is in direct contact with three water molecules, which have penetrated into the pocket and form a layer between Asp26 and the backbone groups to the left (Pro76, Thr77; Figure 5). A fourth water extends this layer and links it to Lys57, which is hydrogen-bonded to Asp26, as discussed. Note that despite these nearby waters, ionized Asp26 is considerably less solvated than the solvent-exposed Asp20, whose surrounding water distribution is also shown in Figure 6.

Relation to Linear Response Theory. $\partial G/\partial\lambda$ is essentially the electrostatic potential on the Asp carboxylate group (see eq 5). Its decrease with λ, shown in Figure 7, describes the increasing polarization of the Asp environment in response to the growing Asp charge. The decrease is approximately linear for Asp in solution (not shown) and for Asp26 in the protein. This indicates that the environment responds approximately as a linear medium; i.e., the polarization density induced by the new charge increases linearly with λ.[61]

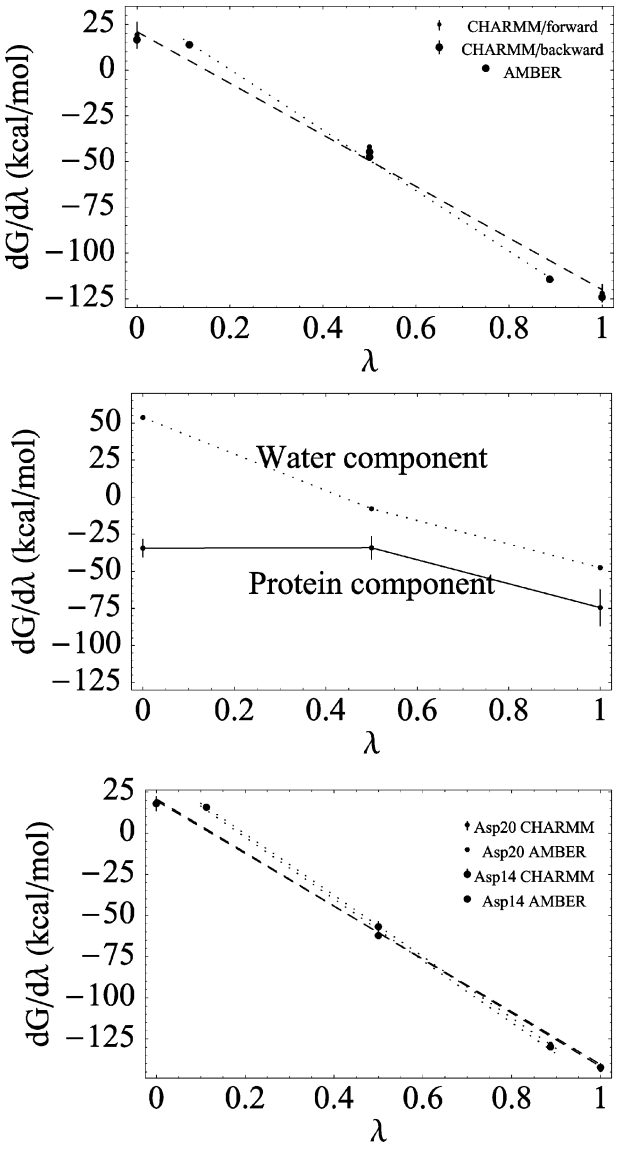

Figure 7. Free energy derivatives as a function of λ. Upper panel: Asp26 runs. Middle panel: Asp26 CHARMM forward run; protein and solvent components of $\partial G/\partial\lambda$. Bottom panel: Asp20, Asp14 runs.

The linearity of $\partial G/\partial\lambda$ is also supported by the approximately Gaussian fluctuations of the energy gap, illustrated in Figure 8. Perfect linearity would imply that the fluctuations are exactly Gaussian, and vice versa.[77] Some deviations from Gaussian statistics are apparent in some of the simulations (Figure 8). The most obvious deviation is for the $\lambda = 0.112\,70$ state, AMBER run, where there are clearly two underlying distributions. They correspond to two distinct populations, or substates, having Asp26 in the inward and outward orientation, respectively. Each of them is roughly Gaussian, although there are significant deviations in the tails of the two Gaussians. The two Gaussians (or parabolas in the log scale of the figure) are centered on the derivative values corresponding to the two substates, −62.0 and −68.4 kcal/mol, respectively (Table 2).

In general, the presence of two substates, each of which has Gaussian statistics, can lead to marked deviations from linear

(77) Levy, R.; Belhadj, M.; Kitchen, D. Gaussian fluctuation formula for electrostatic free energy changes. *J. Chem. Phys.* **1991**, *95*, 3627−3633.

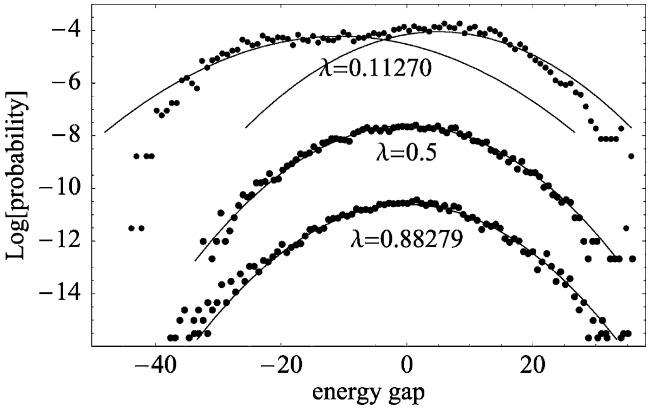

Figure 8. Histograms of the energy gap for different values of the reaction coordinate λ, offset vertically for clarity. Thioredoxin Asp26, AMBER run. For each histogram, a parabolic fit is shown in light gray. For the uppermost panel, two parabolas are used, corresponding to the two substates defined by the Asp26 carboxylate orientation; each substate is only approximately Gaussian, as seen by the deviations in the histogram tails.

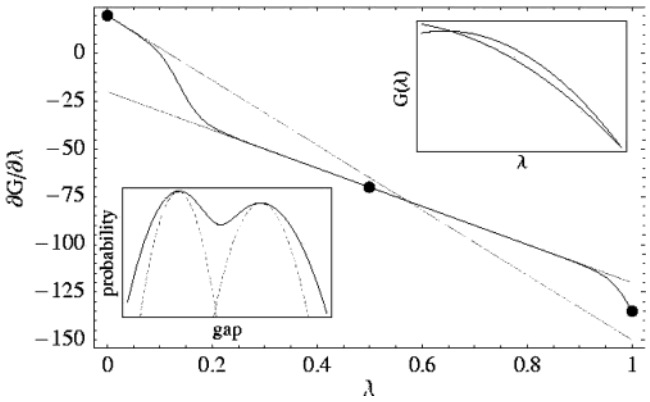

Figure 9. Free energy derivative $\partial G/\partial\lambda$ as a function of λ for a hypothetical system with two Gaussian substates. The derivatives for each individual substate are shown as gray lines; the overall $\partial G/\partial\lambda$ is shown as a black line and seen to be highly nonlinear. Black dots highlight the endpoints and the midpoint. Upper inset: the free energy as a function of λ for each substate. Lower inset: log probability of the energy gap, obtained as a superposition of the Gaussian probabilites of the two substates. The present hypothetical data are roughly compatible with the thioredoxin Asp26 numbers, but for the purposes of illustration, the nonlinearity here is stronger.

response, as illustrated by hypothetical data in Figure 9. The hypothetical data are roughly similar to Asp26, but with a stronger nonlinearity, for illustration. They correspond to a total free energy change of $\Delta G = -70$ kcal/mol and a free energy difference between substates of 5 kcal/mol at the $\lambda = 0$ endpoint; the midpoint derivative is -70 kcal/mol, in exact agreement with ΔG, despite the nonlinearity. The derivative $\partial G/\partial\lambda$ has nearly linear segments in regions where one of the two substates is more stable, with a crossover to the other substate when the substate free energies intersect (see upper inset in Figure 9). Evidently, with only three simulation points (e.g., the dots in Figure 9), care must be taken before inferring linear response.

Free Energy Component Analysis. Following the ideas of electron transfer theory,[4,59,78] it is instructive to decompose the ionization free energy into two components. The first is associated with structural reorganization, and is termed the reorganization, or relaxation free energy; the remaining component is termed the static free energy. This decomposition corresponds to a hypothetical, two-step reaction.[59,78−80] In the first step, the proton is removed, with the system constrained in its initial, reactant state ensemble. In the second step, the system is allowed to reorganize, or relax in response to the ionization. The reorganization free energy, ΔG_{rlx} provides a quantitative measure of the dielectric relaxation occurring in response to the ionization. Since the present system responds approximately as a linear medium, the reorganization free energy is given by the second derivative of the free energy

$$\Delta G_{\mathrm{rlx}} = \frac{1}{2}\frac{\partial^2 G}{\partial \lambda^2} \tag{12}$$

For the model compound Asp in solution, $\Delta G_{\mathrm{rlx}} = -82.1$ kcal/mol with the CHARMM force field and $\Delta G_{\mathrm{rlx}} = -95.7$ kcal/mol with AMBER. For Asp26 in the protein, the values are -71.4 kcal/mol with CHARMM and -83.3 kcal/mol with AMBER. Note that the differences between the protein and model compound are almost the same with the two force fields. The relaxation is much weaker in the protein, as expected, since the protein environment is much less polarizable than bulk water.

For a spherical solute of radius R in a dielectric continuum, the relaxation free energy would be related to the dielectric constant ϵ of the medium by the Born equation[11,25]

$$\Delta G_{\mathrm{rlx}} = \frac{1}{2R}\left(1 - \frac{1}{\epsilon}\right) \tag{13}$$

Taking the model compound relaxation free energy, along with $\epsilon = 80$ (the dielectric constant of water) gives $R = 2.0$ Å with the CHARMM force field and $R = 1.7$ Å with AMBER. These can be viewed as the 'electrostatic size' of the carboxylate group with each force field. The distribution of charge representing the proton is more concentrated with AMBER (Figure 2), giving a larger relaxation free energy and a smaller R. Now, taking the reorganization free energy computed for Asp26 in the protein and using the same effective radii R, we obtain the dielectric constant $\epsilon = 7.1$ (the two force fields giving the same result). This dielectric constant corresponds to a uniform dielectric medium that would produce the same reorganization free energy as the heterogeneous, protein/solvent environment of Asp26. It measures the mean effective polarizability of the environent of Asp26. A similar effective dielectric constant was observed in the active site of the enzymes aspartyl-tRNA synthetase[29,30] and trypsin.[81]

Another useful decomposition separates the free energy into protein and solvent components.[82] Indeed, the free energy derivative takes the form of a pairwise sum of interactions between the Asp26 carboxylate and the rest of the system, which can be split into protein and solvent contributions. The two are plotted as a function of λ in Figure 7 (middle panel), based on the CHARMM, forward run. Both are distinctly nonlinear functions of λ, even though their sum is nearly linear. The protein response does not begin to develop until after the half-

(78) Marcus, R. Chemical and electro-chemical electron transfer theory. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155−196.
(79) Simonson, T.; Perahia, D.; Brünger, A. T. Microscopic theory of the dielectric properties of proteins. *Biophys. J.* **1991**, *59*, 670−90.
(80) Muegge, I.; Qi, P.; Wand, A. J.; Chu, Z.; Warshel, A. Reorganization energy of cytochrome *c* revisited. *J. Phys. Chem. B* **1997**, *101*, 825−836.
(81) King, G.; Lee, F.; Warshel, A. Microscopic simulations of macroscopic dielectric constants of solvated proteins. *J. Chem. Phys.* **1991**, *95*, 4366−4377.
(82) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden thermodynamics of mutant proteins: A molecular dynamics analysis. *Science* **1989**, *244*, 1069−1072.

Table 3. Thioredoxin Asp20 and RNAse A Asp14: Explicit Solvent MDFE Results[a]

	AMBER			CHARMM		
	model	Asp20	Asp14	model	Asp20	Asp14
run length (ns)	3	15	15	6	6	2.8
$\partial G/\partial\lambda$ ($\lambda = 0$)				19.3(1.4)	19.0(0.4)	17.9(4.2)
$\partial G/\partial\lambda$ ($\lambda = 0.11270$)	−1.3(0.8)	−1.0(1.6)	−1.3(1.0)			
$\partial G/\partial\lambda$ ($\lambda = 0.5$)	−75.3(0.2)	−73.9(0.4)	−78.3(1.8)	−58.4(0.8)	−57.0(1.2)	−56.8(3.2)
$\partial G/\partial\lambda$ ($\lambda = 0.88279$)	−148.6(2.0)	−145.1(1.8)	−146.8(3.4)			
$\partial G/\partial\lambda$ ($\lambda = 1$)				−144.6(1.4)	−143.5(0.8)	−142.3(2.4)
ΔG	−75.1(1.1)	−73.4(0.6)	−75.9(1.1)	−60.5(0.6)	−59.6(0.6)	−59.4(2.0)
$\Delta\Delta G$	0	1.7(1.3)	−0.9(1.6)	0	0.9(0.8)	1.1(2.1)
$\Delta\Delta G$ (exper.)		0.0	−2.7		0.0	−2.7

[a] All energies in kcal/mol. Statistical uncertainty in parentheses. For the derivatives, it is estimated as twice the standard deviation of block averages, each trajectory being divided into four blocks. For the free energies, calculated by trapezoidal (CHARMM) or Gaussian (AMBER) integration, simple propagation of error is assumed.

ionized state is reached, once the Asp26-Lys57 salt bridge begins to be populated. Integrating the protein component from $\lambda = 0$ to 1 gives −44.3 kcal/mol, close to the total free energy change, −48.0 kcal/mol. The remaining, solvent component is only −3.7 kcal/mol.

3.2 Asp20 in Thioredoxin. Explicit Solvent Simulations. The free energy data for thioredoxin Asp20 are summarized in Table 3 and in Figure 7 (bottom panel). The experimental pK_a is unshifted, being the same as that of the model compound. Both AMBER and CHARMM predict a small, upward shift. But the shift is only 0.5−1 pK_a units, which is within the statistical uncertainty of the simulations. The free energy derivative is a linear function of the reaction coordinate λ. Asp20 is completely exposed to the solvent (see the radial distribution function, Figure 6), which, indeed, is expected to respond as a linear medium to the ionization reaction.

3.3 Asp14 in Ribonuclease A. Explicit Solvent Simulations. For Asp14 in ribonuclease A, the experimental pK_a is downshifted by 2 pK_a units, corresponding to a double free energy difference $\Delta\Delta G = -2.7$ kcal/mol. The AMBER simulation underestimates the downshift, with $\Delta\Delta G = -0.9$ kcal/mol. The CHARMM simulation predicts a small upshift, $\Delta\Delta G = +1.1$ kcal/mol, with an estimated uncertainty of 2.1 kcal/mol. This is essentially the same as the CHARMM result for thioredoxin Asp20 (above). The simulations were done with all carboxylates ionized except for Asp14, despite the low pK_a. Because these groups are solvent exposed and rather far from Asp14, we hypothesized that this would not affect the calculations strongly. In fact, MDFE calculations with the GB/OBC model were done to test this hypothesis (see below); they indicate that if the carboxylates are protonated, the ionization free energy in the protein decreases by 1.1 kcal/mol, making the ionized state more favorable. If we subtract 1.1 kcal/mol from $\Delta\Delta G$, the AMBER value becomes −2.0 kcal/mol, very close to the experimental result. The CHARMM value becomes 0.0 kcal/mol.

The CHARMM force field was shown recently to underestimate the solvation free energy of the acetic acid group by 3 kcal/mol (B. Roux, personal comunication). This could possibly account for the rather high calculated CHARMM pK_a, although some cancellation of error is expected when the protein and model compound free energy changes are subtracted. Another possibility is that the experimental pK_a is measured at such a low pH that the vicinity of Asp14 is structurally modified or partly denatured. Some evidence for a conformational change occurring in this region at around pH = 5 is given by two-dimensional NMR measurements.[19] However, the good agreement obtained with the AMBER force field suggests that this does not strongly affect the pK_a. The CHARMM error may also be due to the somewhat short run length (2.8 ns), which may not give the ionized state time to stabilize completely. The AMBER run length was 15 ns.

Despite uncertainty in the free energy, the simulations provide several insights into the structure and dynamics of the Asp14 environment. In the crystal structure, determined at a pH of 5.2 (where Asp14 is ionized), Asp14 is in a congested region of the protein surface, with two positively charged side chains nearby, His48 and Arg33, which presumably contribute to the low experimental pK_a. The simulations are in good agreement with the crystal structure, with typical rms deviations of just 1.2−1.8 Å for the protein backbone. In the simulations of the protonated state, the free Asp14 oxygen forms a tight hydrogen bond to the Tyr25 hydroxyl group. In the ionized state, this interaction weakens, but new hydrogen bonds are formed with Thr17, His48, and Ser16. The carboxylate group also interacts with the positively charged Arg33. The average distance between Asp14-OD and Arg33-NH is 4.3 Å in the protonated state; it is only 2.9 Å in the ionized state (taking the OD and NH closest to each other), which corresponds to a salt bridge.

To summarize, the Asp14 downshift calculated with the AMBER force field appears to be due to hydrogen bonds between ionized Asp14 and several nearby polar groups, combined with a salt bridge interaction with Arg33 in the ionized state. Deviations from the experimental pK_a could arise from force field deficiencies. It may be that the mean field treatment of electrostatic interactions employed in the present force fields is not sufficient for such a heterogeneous environment. Thus, even though the conformations sampled may be correct, the charge−charge energetics could be slightly off. Notice that the same limitation is inherent in every continuum electrostatic approach implemented to date.[10,21−24,39−41]

For Asp14, as for Asp20 in thioredoxin, the free energy derivative is a linear function of λ, indicating that the heterogeneous environment responds as a linear medium to the ionization reaction. For both Asp14 and thioredoxin Asp20, the reorganization free energy is about as large as that for the model compound in water, even though a large part of the nearby space is occupied by protein, which is a less polarizable medium.

3.4 Implicit Solvent Simulations. The three pK_a's were also calculated using molecular dynamics with a continuum solvent model. This work represents the first application of MDFE methods to a pK_a calculation with continuum solvent. The generalized Born (GB) model was used to describe the

Table 4. Implicit Solvent MDFE Results[a]

	ΔG		$\Delta\Delta G$		
system	GB/OBC	GB/ACE	GB/OBC	GB/ACE	exp.
model	−65.0(0.2)	−43.1(0.1)	0	0	0
Asp26	−57.9(1.0)	−32.0(0.9)	7.1(1.0)	11.1(0.9)	4.8
Asp20	−64.5(0.4)	−42.5(0.2)	0.5(0.4)	0.6(0.2)	0.0
Asp14	−66.1(5.9)		−1.1(5.9)		−2.7
Asp14[b]	−67.2(2.1)		−2.2(2.1)		−2.7

[a] Free energies in kcal/mol. [b] All carboxylates protonated except Asp14.

continuum solvent efficiently, allowing long simulations (15 ns for each pK_a). Two variants of the GB model were employed, GB/OBC and GB/ACE (see Methods). For both variants, the MD structures are in fair agreement with the crystal structures, with typical backbone rms deviations of 1.5−2.5 Å with both GB/ACE and GB/OBC. The free energy results are summarized in Table 4. With GB/OBC and AMBER, the calculated pK_a's are closer to experiment than to the explicit solvent results above. The RNase A Asp14 pK_a is downshifted by the correct amount (particularly when all the other carboxylates are protonated), and the thioredoxin Asp26 pK_a is only 1.5 units higher than experiment (9 instead of 7.5 experimentally). This is encouraging, since traditional PB methods, which use a single fixed structure, have great difficulties in reproducing such anomalous pK_a's without resorting to adjustment of parameters. The Asp20 pK_a is unshifted, as it should be. With GB/ACE, the Asp20 and Asp26 double free energy differences are $\Delta\Delta G = 0.6$ and 11.1 kcal/mol, respectively, in close agreement with the explicit solvent results (Tables 1, 3), even though the protein force fields are different in the two cases (CHARMM/param19 here, vs CHARMM/param22 with explicit solvent). The Asp14 pK_a was not computed with GB/ACE.

Interestingly, with GB/ACE, when Asp26 becomes ionized, its environment does not undergo the same relaxation as with explicit solvent. Thus, Lys57 never approaches Asp26 more closely than 5.7 Å, and the Asp26-Lys57 distance is essentially independent of the reaction coordinate λ. The Asp26-Lys57 salt bridge is replaced by interactions between Asp26 and the continuum solvent. Despite this tradeoff between protein and solvent interactions, the pK_a's with explicit solvent and GB/ACE are in very good agreement. Another example of an 'equal free energy' tradeoff between a salt bridge and solvent-mediated interactions was found recently in simulations of an enzyme active site.[83] With the AMBER force field and GB/OBC, the Asp26-Lys57 salt bridge forms as before.

4. Discussion

MDFE with Explicit Solvent: Limitations and Difficulties. The first application of MD free energy simulations to proteins was a pK_a shift calculation.[31] Encouraging results were obtained, and the authors concluded that 'quantitative results for the energetics of solvated proteins' were finally at hand. Since then, there is no question that a large and important body of data has been obtained from theoretical studies of protein electrostatics. Still, almost twenty years later, the difficulty to compute very accurate pK_a shifts is striking. The interesting cases are also the difficult ones, where the titrating group is in a congested and/or partly buried environment, so that individual, ordered water molecules are important, and rare, activated protein motions participate in the dielectric relaxation. In a case like thioredoxin Asp26, the structural reorganization involves concerted motions of protein backbone and side chain groups, as well as ordered waters, which take place on nanosecond time scales.

Very long simulations were employed here with two different force fields, in an attempt to clearly identify and sample the important degrees of freedom while directly measuring dependency on model details. The structures sampled were stable and in good overall agreement with the crystal structures (e.g., Figure 3, bottom panel). However, even with 35 ns of total run length for Asp26, the apparent statistical uncertainty is more than one pK_a unit. For the two difficult cases, Asp26 (thioredoxin) and Asp14 (RNase A), the AMBER/CHARMM differences are 2−3 kcal/mol, slightly more than the statistical uncertainty. This is reassuringly moderate, but certainly not negligible.

The deviation from experiment for thioredoxin Asp26 and RNase A Asp14 is greater than the statistical error; for Asp26 it is twice as large as the force field difference. Both of these pK_a's are too high with both force fields. With CHARMM, this may be due in part to undersolvation of the protonated, acetic acid group in the CHARMM22 force field (B Roux, personal communication); i.e., the ionized model compound may be computed as too stable relative to the protonated model compound. The exact extent of this effect is difficult to determine for the protein leg of the thermodynamic cycle (Figure 1). The CHARMM run lengths may also be too short, so that the stability of the final, ionized state is underestimated. Thus, the CHARMM Asp14 run length was only 2.8 ns, compared to 15 ns with AMBER. The AMBER pK_a's are also too high, though not as high as with CHARMM. For Asp26, this could arise from an Asp26-Lys57 interaction that is too weak, possibly because many body induction effects (electronic polarizability) are underestimated for this site by the mean field approach. Polarization of the Asp26 carboxylate by the lysine charge contributes to their interaction; this effect is presumably stronger and more anistropic in the largely buried thioredoxin site than in a typical exposed region, and may be underestimated by both force fields.

Structural and Free Energy Component Analyses. Even though the pK_a estimates are not perfect, the simulations with explicit solvent yield a great deal of structural and thermodynamic information. The contributions of protein and solvent to the free energy were calculated for Asp26, and the most important structural groups were identified. The two possible protonation sites on the Asp26 side chain were compared in detail. With the AMBER force field, the protonation free energy is only mildly dependent on the orientation of the carboxylate. With the CHARMM force field, one orientation is clearly preferred. Most Poisson−Boltzmann calculations neglect this preference (although it can in principle be included in the methodology[39−41]).

The magnitude of dielectric relaxation in response to ionization was characterized for all three aspartate residues, and the dielectric response was shown to be approximately linear, even though the protein and solvent contributions taken separately can be distinctly nonlinear. For Asp26, the different proton affinities of the 'inner' and 'outer' carboxylate oxygens (with the AMBER force field) lead to two distinct conformational

(83) Archontis, G.; Simonson, T.; Moras, D.; Karplus, M. Specific amino acid recognition by aspartyl-tRNA synthetase studied by free energy simulations. *J. Mol. Biol.* **1998**, *275*, 823−846.

substates and to a moderate nonlinearity in the dielectric response. A mean polarizability was estimated for the Asp26 environment from the Born equation. It corresponds to a mean dielectric constant of 7.1, a value intermediate between water and a nonpolar, alkane-like medium, and similar to other enzyme active sites.[29,30,81] For Asp20 and RNase A Asp14, the reorganization free energy is as large as for the model compound in bulk water, even though much of the volume around these residues is occupied by protein rather than high-dielectric water.

In the case of Asp26, structural reorganization in response to ionization involves the formation of a salt bridge, with Lys57 being drawn into the Asp26 pocket, away from solvent. The unfavorable partial desolvation of Lys57 partly compensates for its favorable interaction with Asp26. At the same time, a few water molecules squeeze into the pocket between Asp26 and the backbone moieties of Pro76 and Thr77. Most of the reorganization free energy arises from solvent (including bulk solvent).

In the case of Asp14, ionization strengthens hydrogen bonds with three nearby polar side chains and leads to a salt-bridge interaction with Arg33. The solvent free energy component was not computed separately, but from the overall magnitude of the reorganization free energy, we conclude that most of it is contributed by solvent.

Comparison to Earlier Poisson−Boltzmann Studies. Two research groups have calculated the pK_a of Asp26 in oxidized thioredoxin by solving the Poisson−Boltzmann (PB) equation, using ensembles of structures obtained from X-ray or NMR experiments.[26,27] Both groups proposed that desolvation of Asp26 and interactions with the positively charged Lys57 are the main factors affecting the Asp26 pK_a. However, the two groups obtained different results based on similar structures. Bashford et al.[27] used two sets of structures with varying distances between Asp26 and Lys57. They found agreement with the experimental pK_a with a set of so-called 'loose' structures, where the Asp26(CG)-Lys57(NZ) distance is on average 8.1 Å. A second set of 'tight' structures, where the distance was ∼4.8 Å, led to a considerably lower value, pK_a = 2.2 on average. Sharp and collaborators[26] also studied the dependence of the pK_a value on the distance between Asp26 and Lys57. They found, contrary to Bashford, that agreement could only be obtained when the side chain dihedrals were rotated to bring Lys57 into salt bridge contact with Asp26 ('tight' structure, ∼4.0 Å separation). Both of these PB studies found that even very small changes in the Asp26-Lys57 distance could have a large impact on the pK_a value. A change of just 0.25 Å could change the pK_a by more than one unit. This underlines how sensitive electrostatic interactions are to conformational fluctuations and how important it is to include these in the calculations. Our MDFE calculations with explicit solvent show that both the 'loose' and the 'tight' structures are in fact relevant, with the protein switching between them when Asp26 becomes ionized. They also suggest that nonelectrostatic interactions play a role, for example through the increased burial of the alkane portion of the Lys57 side chain. PB calculations based only on 'tight' or 'loose' structures do not allow explicit relaxation of Lys57; instead, they must mimic its relaxation by a rearrangement of polarization charge on the protein−solvent interface. This description is too crude to make an accurate prediction of the pK_a (as pointed out by Bashford et al.),[27] even though the experimental pK_a can be reproduced by tuning the dielectric constant.

Methods to include explicit protein relaxation in a PB pK_a calculation do exist. The simplest one is likely to be sufficient in many cases. However, it requires structural models for both endpoints of the reaction. Since the PB model is linear, it is then sufficient to compute the static free energy for deprotonation in each state and take the average over the two states. The static free energy is the electrostatic potential times the charge increments, and is readily available with PB. This method has been advocated by several groups[16,28−30] using somewhat different terminologies, but is not widely used because both endpoint structures are needed. Results for the present systems with this method will be presented elsewhere. Another approach is to include several protein conformations in the PB model and allow their occupancies to vary as a function of the titration state.[39−41] This method gives much better results than the standard, single conformation approach.

First Example of MDFE with Dielectric Continuum Solvent. Another simplified model treats the protein and its fluctuations in atomic detail, while treating the solvent as a dielectric continuum. This work represents the first MD free energy simulation of a protein with a dielectric continuum solvent. A related, earlier approach employed MD with explicit solvent to sample conformations of the ionized and neutral protein states, then used a Langevin Dipole solvent model to evaluate their energetics.[28] Results with a similar method will be presented elsewhere for thioredoxin and RNAse A (Archontis, JC, DAC & TS; in preparation). Here, the same continuum solvent is used for the conformational sampling and the free energy estimation. This is simpler and more efficient, but the conformations sampled are probably less accurate (see below). The Generalized Born continuum solvent model was used, allowing 15 ns of MD for each pK_a, leading to good convergence. MDFE using numerical solutions of the PB equation is possible in theory, but still too slow to be practical.[44,45,84]

The GB results for Asp26 are close to the explicit solvent ones, even though the structural relaxation differs in the two cases. In particular, with GB/ACE, the Asp26-Lys57 salt bridge does not form in the ionized state, presumably because the balance between Lys57 desolvation and Asp26-Lys57 interaction has changed, compared to the explicit solvent model. Because the overall quality of the structures sampled with explicit solvent is superior for this system and others,[69] we believe the explicit solvent simulations give the correct picture, with formation of the salt bridge in the ionized state. But the 'loose' conformations are probably not very much higher in free energy, and indeed, the AMBER simulations of the almost-ionized state (λ = 0.882 79) spent about half of the time in a 'loose' conformation.

Overall, with the AMBER force field and the GB/OBC variant of the GB model,[70] the pK_a results are actually better than with explicit solvent, and the rms deviation from experiment is just one pK_a unit. With the CHARMM19 force field and the GB/ACE variant, the pK_a results are very close to the explicit solvent results (with the CHARMM22 force field). This suggests that the deviations from experiment come mainly from limitations of the protein force field, which can be corrected

(84) David, L.; Luo, R.; Gilson, M. Comparison of Generalized Born and Poisson models: energetics and dynamics of HIV protease. *J. Comput. Chem.* **2000**, *21*, 295−309.

systematically, and to a lesser extent from the solvent model. The structures are in reasonable agreement with experiment. Although the GB variants used here are better than some earlier ones, more recent, improved variants already exist,[85] and there is room for further improvement, leading to even better structures and energetics. This should open the way for large-scale titration studies, where many amino acids are allowed to bind and release protons during the course of an MD simulation.[86,87]

5. Conclusions

Several computational approaches have been used to study pK_a's of side chains in proteins. These have typically employed continuum models to account for the dielectric response of the solvent, and have employed various methods to model the energetics and to account for conformational reorganization in the protein.

Somewhat surpisingly, aside from pioneering early work[31] and one recent study,[28] this problem has not been addressed by combining modern simulation techniques (that consider the protein and explicit solvent) and well-established free-energy methods such as thermodynamic integration. In principle, such free energy simulations can yield results that are as accurate as the underlying force fields themselves, and recent results on the analogous oxidation/reduction problem[61] and a proton-transfer problem[88] have been very promising. Here, we show that molecular dynamics free energy (MDFE) methods can indeed give useful results for pK_a's of carboxylic acid side chains in proteins, but that care is needed to ensure convergence of the results and (in some cases) to avoid conformational drift away from the most relevant solution conformations.

The key advantage of the procedure outlined here is that (in principle) it correctly computes the required protonation free energies, including both entropic and enthalpic effects. The procedure incorporates any conformational changes that may accompany protonation, and models the molecular details of water solvation. Long-range electrostatic effects, which can be quite important for charging free energies, can be incorporated with Ewald or reaction-field methods that eliminate the need to impose any cutoffs for Coulombic interactions. For a given force field, such results can then become benchmarks against which more approximate calculations can be compared. By modeling at least the end points of the protonation events in full detail, we can discover which features most affect the thermodynamics, and can better assess the strengths and weaknesses of current force fields and simulation methods.

It is certainly true, however, that MDFE simulations are neither necessary nor desirable for all problems, and we list here some of the disadvantages and limitations of this approach.

1. The rate at which the equilibrium ensemble is reached from particular starting configurations is not known in advance, and may be quite slow. This can be especially true for partially buried residues (such as Asp26 in thioredoxin), where "interior" water solvation is, on average, significantly different in the protonated and de-protonated forms. Simulations of multiple nanoseconds can be required, inhibiting the routine application of this model to large numbers of sites.

2. Long simulations often have a tendency for the overall protein structure to slowly drift away from conformations derived from X-ray or NMR experiments. This drift presumably reflects inadequacies in current force fields, and seems to be more pronounced for generalized Born simulations than for those with explicit solvent. When this happens, the accuracy of the results can actually decline with longer simulations, since early configurations sample the correct conformational ensemble more closely than do later ones.

3. The results are still somewhat sensitive to the force field that is used, or to the use of explicit vs implicit solvation models. This is of course a feature of all simulations, but it may be that simpler models that are empirically tuned to reproduce protein pK_a values might perform better (with respect to experiment) than the more rigorous approaches used here.

4. Proteins generally have multiple side-chains that titrate in a given pH range, and it is not straightforward to include such site−site interactions into the framework of detailed MDFE calculations. Some models, such as grand-canonical simulations (in which protons can transfer between the system being modeled and a hypothetical bath at a given pH) appear to be promising,[86,87] but much simpler models of the energetics (dating back to Linderstrom-Lang, Tanford, and Kirkwood) may currently be more appropriate for problems in which the global nature of pH behavior, rather than individual residues, is emphasized.[21,39,41]

Despite these limitations, the prospects of MDFE simulations of protein pH behavior seem very strong. Conformational changes, whether large or small, are always coupled to protonation events, and methods that rely upon X-ray or NMR structures corresponding to just one protonation state are not sufficient. Even relatively small conformational readjustments have important thermodynamic consequences that in the end can only be understood from models that have a rigorous statistical mechanical foundation. In this work, using MDFE, the existence and direction of the pK_a shifts were correctly predicted for three aspartates with the AMBER force field, and in two out of three cases with the CHARMM force field. A wealth of structural and thermodynamic information was obtained, providing detailed insights into mechanisms of pK_a shifts in proteins. The complexity and the importance of conformational reorganization was demonstrated, and deviations from a simple linear response were observed for one of the three aspartates. The role of nonelectrostatic interactions was emphasized. The polarizability of the thioredoxin active site was characterized roughly by a mean dielectric constant of about 7, similar to other enzyme active sites that have been analyzed this way. Finally, encouraging results were obtained by treating the protein dynamics explicitly while treating the solvent as a dielectric continuum. This indicates that although individual, nearby waters reorganize strongly around the titrating groups, it is the long-range electrostatic properties that dominate the contribution of solvent to the thermodynamics.

Acknowledgment. We acknowledge financial support from the Centre National de la Recherche Scientifique and the

(85) Lee, M.; Salsbury Jr., F.; Brooks, III, C. Novel generalized Born methods. *J. Chem. Phys.* **2002**, *116*, 10 606−10 614.
(86) Börjesson, Ü.; Hünenberger, P. H. Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. *J. Chem. Phys.* **2001**, *114*, 9706−9719.
(87) Baptista, A.; Martel, P.; Petersen, S. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. *Proteins* **1997**, *27*, 523−544.
(88) Schweins, T.; Geyer, M.; Scheffzek, K.; Warshel, A.; Kalbitzer, H.; Wittinghofer, A. Evaluation of catalytic free energies in genetically modified proteins. *Nat. Struct. Biol.* **1995**, *2*, 36−44.

National Science Foundation (CNRS/NSF travel grant INT-0128783; to D.A.C. and T.S.), from the French Ministère de la Recherche (Program inter-EPST de Bioinformatique; to T.S.), and from the National Institutes of Health (Grant NIH GM57513; to D.A.C.), as well as helpful discussions with Charles L. Brooks III and Georgios Archontis. Some of the calculations were performed at the supercomputer center of the French Ministère de la Recherche (CINES; supercomputer allocation to T.S.).

JA039788M