A Poisson−Boltzmann Study of Charge Insertion in an Enzyme Active Site: The Effect of Dielectric Relaxation

Thomas Simonson,*,† Georgios Archontis,‡,§ and Martin Karplus*,‡,||

Laboratoire de Biologie Structurale (C.N.R.S.), I.G.B.M.C., 1 rue Laurent Fries, 67404 Illkirch (C.U. de Strasbourg), France, Laboratoire de Chimie Biophysique, Institut Le Bel, 4 rue Blaise Pascal, Université Louis Pasteur, 67000 Strasbourg, France, Department of Natural Sciences, University of Cyprus, P.O. Box 537, CY 1678, Nicosia, Cyprus, and Department of Chemistry & Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received: April 26, 1999

Continuum solvent models are playing an increasing role in the study of aqueous solutions, particularly those involving protein solutes. To estimate the magnitude of dielectric relaxation and clarify the microscopic meaning of the protein dielectric constant, charge insertion in the active site of the enzyme aspartyl-tRNA synthetase (AspRS) is analyzed using finite-difference Poisson−Boltzmann calculations. The insertion process is a simplified model that mimics qualitatively the mutation of substrate Asp into Asn, studied earlier by free energy simulations. A two-step insertion path gives the relaxation and nonrelaxation ("static") free energy components separately. The assumption of linear response leads to a linear relation between the two components, connecting the explicit structural differences between reactant and product structures with the relaxation free energy calculated from either structure. This relation is verified here only if protein dielectric constants of 1 and 4−8 are used for the static and relaxation free energies, respectively. These are also the only conditions that give reasonable agreement with the Asp → Asn free energy simulated earlier and with a molecular dynamics/linear response estimate of the present charging free energy. The use of two protein dielectric constants represents a significant departure from standard continuum models. The values obtained are physically reasonable: a protein dielectric of 4−8 for relaxation indicates that the active site of AspRS, though highly polar, is only moderately polarizable. A dielectric of one for the static term indicates that the charge set, optimized for explicit solvent simulations, reproduces the equilibrium potential without the need for additional, implicit protein polarization. In contrast to simple charge insertion, the binding free energies of Asp and Asn to AspRS are best calculated with a more standard protocol that uses a single protein dielectric of 4, accurately reproducing free energy simulation results. For this and other binding processes, additional free energy components are involved, related to desolvation of the binding site; the optimal dielectric constant represents an empirical compromise among these. A multistep component analysis could also be used to analyze the role of relaxation in these more complex processes. It is suggested that the use of more than one dielectric constant in continuum models will lead to a more consistent and robust description of dielectric phenomena in solution.

1. Introduction

The goal of this study is to analyze dielectric relaxation in an enzyme active site with the help of a continuum model. Although continuum solvation models have been used successfully to examine many problems that are dominated by electrostatic effects,[1,2] there are several questions that arise in their application. The first of these concerns the dielectric constants or dielectric functions to be used for the solute; for the aqueous solvent the experimental value of ∼80 at room temperature is appropriate, except for the treatment of solvent molecules that interact strongly with the solute. Values ranging from 1 to 20 have been employed for solutes, depending on their nature (e.g., small molecules or proteins) and the type of process being considered (e.g., pK_a's or ligand binding). Although there have been attempts to justify the chosen values,[3,4] it is generally agreed that the solute dielectric constant is still an empirical quantity.[5] More subtle, though related questions concern the possibility that, in fact, it may not be appropriate to use a single dielectric constant to describe macromolecular solutes. The dielectric constant may not be constant throughout the macromolecule; for example, that for the interior may be significantly smaller than that near the surface.[6,7] Also, different dielectric constants may be necessary to describe different responses of the medium, such as electronic relaxation or relaxation involving the motion of charged or polar particles.[8] It is, of course, the latter that makes the dominant contribution to the large dielectric constant of water. In many cases, the structure of the solute changes during the process of interest; e.g., in protein pK_a calculations, the protonated and unprotonated side chain may have different positions, and in ligand binding to proteins, two related ligands whose binding energies are being compared may have different positions relative to the protein. There is the additional question of the time dependence of the phenomena under consideration; i.e., for equilibrium properties the static dielectric constant is appropriate but for dynamic

* Corresponding author.
† Laboratoire de Biologie Structurale.
‡ Laboratoire de Chimie Biophysique.
§ Department of Natural Sciences.
|| Department of Chemistry & Chemical Biology.

10.1021/jp991354j CCC: $18.00 © 1999 American Chemical Society
Published on Web 07/07/1999

processes a frequency-dependent dielectric constant is, in principle, required.[9]

The range of phenomena that may have to be encompassed in continuum electrostatic calculations makes clear why the most common approach is to introduce an empirical dielectric constant that fits some of the data and to use it for related processes. Although such an approach is simple and useful in many cases, it has the inherent difficulty that without an understanding of the phenomenology, the range of applicability of the particular choice of dielectric constant is uncertain. Moreover, a deeper knowledge of the dielectric phenomena in macromolecular systems is of intrinsic interest and should serve as a useful guide to selection of the "correct" simple continuum model.

In the present paper, we focus on the simplest electrostatic process in a protein: insertion of a charge, or a small group of charges, onto existing protein atoms. We employ two approaches to the problems outlined above. In the first, we use molecular dynamics simulations to explore the nature of the electrostatic interactions involved and to determine the continuum model required to interpret the results. The present analysis differs from simple fits to experimental data in that a molecular dynamics simulation provides a detailed description of the reaction being considered, and therefore makes possible a more fundamental analysis of the various contributions to the overall effect. In the second, we introduce certain consistency conditions that are satisfied when the continuum model is exact, and we use them to test the choice of dielectric constants.

A charge binding to a protein interacts with a complex set of charged and polar groups, some of which are essentially fixed in position, while others can move extensively in response to the perturbing charge. In particular, for an ion of finite size, water molecules must be displaced from the binding site to make room for the ion. Such a displacement cannot be described as a simple electrostatic process. However, in the case of a small charge, such as that of a proton or an electron, it is common for continuum models to neglect its volume, and view it as an additional charge placed on existing protein atoms; such models often work well. One can then view the average potential at the perturbing charge position as a superposition of the potential that existed prior to its introduction, and the reaction potential it induces. The former is a measure of the equilibrium polarity of the site in the absence of the charge, while the latter measures the dielectric relaxation of the site and its environment in the presence of charge. This decomposition corresponds to a two-step charge insertion process commonly invoked in electron transfer theory:[10] first, the charge is inserted with the protein and solvent constrained in their unperturbed state; second, the constraints are released and the environment relaxes, or "reorganizes". The corresponding free energies will be referred to below as the "static" and relaxation free energies, respectively.[10−13]

In the context of a continuum model, the distinction between static and relaxation free energies can be very fruitful. In a typical continuum model, the protein contains a set of "permanent" charges located on specific protein groups.[1,2] These in turn give rise to induced volume charge located on the permanent charges, induced surface charge at the protein−solvent interface, and counterion charge in the surrounding solvent. In the first step of charge insertion, all the other charges (permanent and induced) are held fixed. The static free energy is qV_0, where q is the inserted charge and V_0 is the equilibrium potential at the insertion site in the absence of q. V_0 depends in an essential way on the protein permanent charges, as well as on its dielectric constant. In the second step, relaxation occurs. In most continuum models, the protein permanent charges are assumed fixed, so that dielectric relaxation takes the simple form of a redistribution of induced charge. Due to the linearity of continuum electrostatics (in the limit of low ionic strength), the relaxation free energy does not depend on the permanent charges, but only on the shape and dielectric constant of the protein. The relaxation free energy is therefore identical to the self-energy of the inserted charge, as defined in Born theory.[7]

The two charge insertion steps thus correspond to qualitatively different processes, which are sensitive to different properties of the environment. Therefore, as pointed out by Krishtalik and others,[8,14−16] it may be desirable to use different values of the protein dielectric constant for each step. Indeed, a protein dielectric constant of one or two is usually thought to be optimal to calculate the equilibrium potential in a protein (as in the static step),[1,2] at least when a detailed set of protein permanent charges is used (such as a molecular mechanics set).[17] In contrast, the relaxation step may require a larger value: dielectric dispersion by polyamide crystals and dry protein powders suggests that values of around four may be required.[18−20] Molecular dynamics simulations show that the Fröhlich-Kirkwood dielectric constant within the bulk of six proteins is in the range 2−4 (provided charged surface side chains are viewed as part of the solvent medium).[6,15,16] Consequently, by treating the static and relaxation free energies differently, we expect a more accurate continuum model, as well as increased insight into the magnitude and role of each free energy component.

In the analysis of the dielectric constants appropriate for charge insertion, we consider a process that is designed to mimic qualitatively the transformation of substrate Asp into Asn in the active site of the enzyme aspartyl−tRNA synthetase (AspRS). We use this system because results for it are available from molecular dynamics free energy simulations (MDFE).[21,22] Comparing continuum electrostatic potentials and free energies with the molecular dynamics results is used to aid in determining the optimal protein dielectric constants.

The choice of optimal protein dielectric constant(s) is also evaluated by applying an internal consistency condition for the continuum model. This condition is obtained by requiring that the total free energy change be robust with respect to the pathway used for the charge insertion. For the present system, only one combination of protein dielectric constants satisfies this condition; namely, the dielectric constant must be approximately one for the static term and approximately four to eight for the relaxation term. This is also the only combination that gives reasonable agreement with the MDFE result for ΔG(Asp→Asn), and with a molecular dynamics estimate of the charging free energy based on a linear response approximation. The use of two protein dielectric constants represents a significant departure from standard applications of the continuum model. The first value (static free energy) is consistent with our expectation that a low dielectric constant is optimal to calculate the equilibrium potential, given the molecular mechanics charges employed. The second value (relaxation free energy) is consistent with the moderate polarizability of the protein around its active site, estimated separately from molecular dynamics with a linear response approximation (MDLR).

The charge insertion considered here is a particularly simple process, which makes possible a full analysis. When applying continuum models to more complex situations, other free energy components may be present. Thus, when a large ligand binds to a protein, cavity formation is involved, as mentioned above. Nevertheless, the present analysis should be directly relevant to many biochemical processes that can be modeled as an addition or rearrangement of charges on a few existing protein

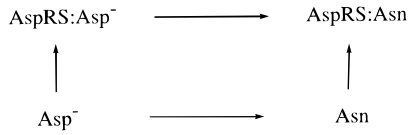

Figure 1. Thermodynamic cycle describing the Asp → Asn mutation in protein (upper horizontal leg) and in solution (lower horizontal leg). Vertical legs correspond to protein−ligand binding.

atoms, such as electron and proton transfer, photoexcitation of bound chromophores, and the binding of small ions. However, it should be noted that although the approach used here is rather general, the specific values of the dielectric constants, particularly that involved in the relaxation free energy, are likely to vary with the system being considered.

In addition to charge insertion in the protein, several related problems have been considered. To obtain the difference in binding free energies between Asp and Asn with the thermodynamic cycle in Figure 1, the Asp → Asn transformation was carried out earlier by MDFE both in the protein and in solution;[22] for comparison, the solution transformation has been carried out using a continuum model. The vertical steps of the thermodynamic cycle represent the actual binding of Asp and Asn to AspRS; such binding steps are commonly modeled with continuum calculations. We have performed such calculations to clarify the relation between the novel two-dielectric protocol used here for charge insertion and more standard continuum protocols commonly used for ligand binding. The calculations and their results will be described elsewhere.

Related problems have been considered recently by several authors. Warshel and co-workers have emphasized the importance of treating dielectric relaxation specifically, e.g., in titration or redox potential calculations,[5,14,23] and discussed at length possible definitions of the protein dielectric constant; they also demonstrated the importance of relaxation free energies in response to charge pair creation in proteins.[24] Several authors have used linear response methods to study charge insertion in proteins.[7,14,25]

The next two sections describe the Theory and Computational Methods. Results are presented in the following section. The last section is the Discussion.

2. Theory

2.1. Distinguishing Static and Relaxation Free Energy. For completeness, we recall here some general aspects of dielectric relaxation theory. We limit ourselves to low ionic strengths and the linearized form of the Poisson−Boltzmann equation. The free energy to insert a charge q into the protein complex is obtained by a thermodynamic integration procedure in which the charge is gradually inserted in infinitesimal fractions. The charge is assumed to be initially in a vacuum (however, see section 3.2 below for a more complicated case). The free energy to increment the charge from λq to $(\lambda + \delta\lambda)q$ is (see, e.g., eq 10.2 of ref 26)

$$\delta G = V(\lambda) q \delta\lambda \tag{1}$$

where $0 \le \lambda < 1$ and $V(\lambda)$ is the potential at the site of charge insertion when the fractional charge λq is present. The latter potential can be written

$$V(\lambda) = V_0 + V^{\mathrm{r}}_{\lambda q} = V_0 + \lambda V^{\mathrm{r}}_q \tag{2}$$

where V_0 is the potential in the absence of q, $V^{\mathrm{r}}_{\lambda q}$ is the reaction potential produced by the fractional charge λq, V^{r}_q is the reaction potential produced by the entire charge q, and the last equality is a consequence of the assumed linear response of the medium. Because the Poisson−Boltzmann equation has been linearized, V_0 does not depend on q, and conversely, V^{r}_q does not depend on the permanent charges other than q. Integrating eq 1 from $\lambda = 0$ to 1 gives the free energy for charge insertion:

$$\Delta G(0 \rightarrow q) = qV_0 + {}^1/_2 qV^{\mathrm{r}}_q \tag{3}$$

The first term in $\Delta G(0 \rightarrow q)$ is linear with respect to q, while the second is quadratic, since V^{r}_q is proportional to q. From linear response theory, it is known that the relaxation free energy in response to a vanishingly small perturbation is quadratic with respect to the perturbing field (or charge), while the remainder of the free energy is linear.[11] Thus, the second term can be identified with the relaxation free energy. It is also easy to prove this explicitly by considering the two-step charging pathway in Figure 2.[7] Generalizing to a group of perturbing charges (such as the Asp side chain carboxylate) is straightforward. Because V^{r}_q does not depend on the permanent charges other than q, it can be calculated equally well in their absence, i.e., the relaxation free energy in response to q is identical to the "self-

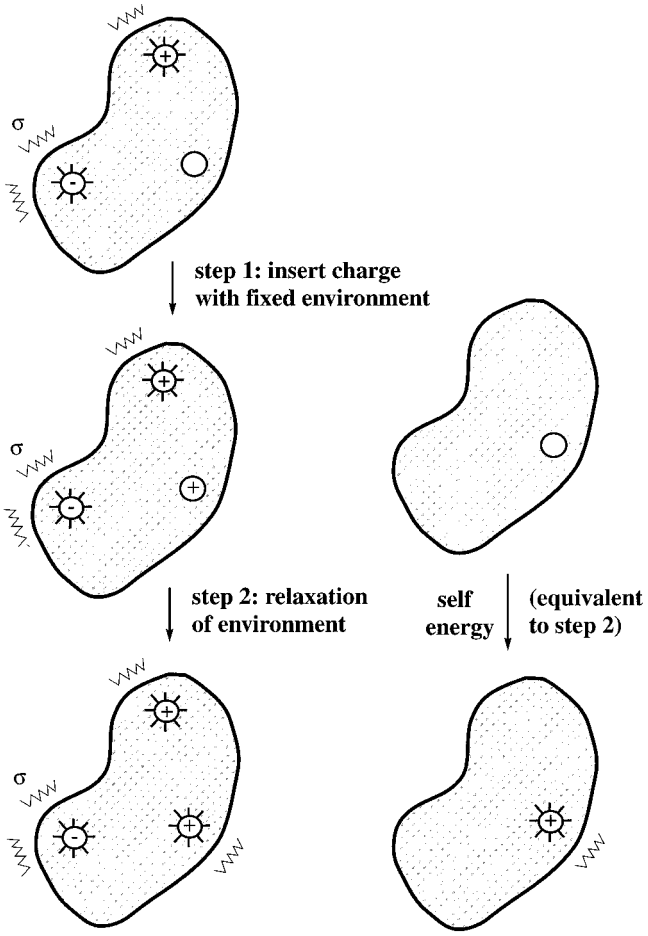

Figure 2. Two-step process for introducing a perturbing charge into a protein. Step 1 gives the static free energy, step 2 the relaxation free energy. The vertical leg on the right gives the charge self-energy, which is equivalent to the relaxation free energy of step 2 (see text). Permanent charges are shown as circles; induced volume and surface charge are schematized respectively as "prickles" around the permanent charges and as "squiggles" at the protein surface. Although each permanent charge gives rise to induced charge spread over the entire protein surface, the surface charge σ is schematically shown near the permanent charges, where it is largest. A counterion cloud can be present but is not shown.

energy" of q.[7] V_0 contains contributions from all permanent charges other than q, and from induced charge produced by those permanent charges. This last point is important because it implies that, in practical calculations, V_0 depends on the protein dielectric constant, even though it does not involve relaxation of the protein in response to the inserted charge.

The relaxation free energy is expected to depend primarily on the polarizability of the region around the ligand. The dominant dielectric relaxation mechanism is expected to be reorientation of polar groups, both in the protein and the solvent.[7,9,27] In the language of linear response theory, the relaxation free energy can be viewed as the product between a (real-space) susceptibility and the square of the perturbing field.[7,11−13,51] It is thus appropriate to consider the protein dielectric constant in the second step of Figure 2 as purely a measure of the relaxation properties, or polarizability of the protein. This is also the meaning of the dielectric constant calculated from linear response theory.[9] Protein dielectric constant calculations from simulations[6,15,16,28−31] suggest that a value significantly greater than one is appropriate for the relaxation free energy. Indeed, to a first approximation, the microscopic fluctuations of at least six proteins are compatible with those of a simple homogeneous dielectric medium with an average dielectric constant of about 3−4.[15] However, a detailed comparison between the microscopic fluctuations and the predictions of dielectric theory indicates that the dielectric constant is not the same for the whole protein. A lower dielectric constant of 1−2 was obtained for the inner half of the proteins studied, while a much larger dielectric constant of 20−40 was obtained for the outermost regions. In the innermost region, electronic polarizability makes the largest contribution to the dielectric constant; in intermediate regions, reorientation of polar groups is predominant. In the outermost region, the large dielectric constant is due entirely to reorientation of charged side chains, which could conceivably be treated as part of the solvent medium.[6,32] As pointed out above, the only permanent charge in the relaxation free energy calculation is the perturbing charge on the ligand side chain, and protein groups are present only to make up the dielectric cavity. Thus, viewing charged protein side chains as part of the high dielectric solvent only changes the size of the low dielectric medium.

2.2. Single and Multiple-Step Insertion Pathways. If structures are available for both the reactant and product states (as here, from prior molecular dynamics simulations), the charge insertion can be done along several pathways. The pathways considered below, for example, are (1) insertion of a full positive charge into the reactant state, (2) insertion of a full negative charge into the product state, (3) insertion of fractional charges into either endpoint state (e.g., a $+\,^1/_2$ charge into the reactant state and a $-^1/_2$ charge into the product state), (4) insertion of fractional charges into both the endpoints and the midpoint (for which structures are also available in the present case). The pathways are schematized in Figure 3. The use of two or more incremental steps to insert the charge q should not be confused with the use of a two-step pathway to insert each fractional charge (Figure 2). Thus, if fractional charges are inserted at several points, it is understood that at each point, the two-step pathway of Figure 2 is used for the insertion, i.e., first insertion into a fixed environment and then relaxation of the environment.

While the static and relaxation free energy components are expected to be pathway dependent, the total free energy should not be, in principle. This leads to a powerful consistency test for the continuum model.

To show how to combine different charging schemes for a single perturbing charge q, we consider the free energy to introduce the fraction λ ($0 \le \lambda \le 1$) of q. It has the form

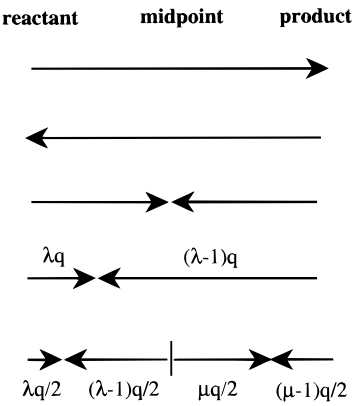

Figure 3. Possible step combinations connecting the Asp:AspRS (reactant) state to the Asn:AspRS (product) state, using one or both endpoint states and possibly the midpoint as starting structures.

$$\Delta G(0\rightarrow\lambda q) = \lambda q \Delta G_s^{reac} + \lambda^2 q^2 \Delta G_r^{reac} \quad (4)$$

where ΔG_s^{reac} and ΔG_r^{reac} are the static and relaxation free energies to introduce a unit charge $q_u = 1$ into the reactant state. With the notations of eq 3, we have $\Delta G_s^{reac} = q_u V_0$, $\Delta G_r^{reac} = q_u^2/2V_1^r$. The free energy to introduce $(\lambda - 1)q$ in the product state is

$$\Delta G(q\rightarrow\lambda q) = (\lambda - 1)q \Delta G_s^{prod} + (\lambda - 1)^2 q^2 \Delta G_r^{prod} \quad (5)$$

where the superscript "prod" indicates the free energies to introduce a unit charge q_u into the product state. Subtracting the two above free energies yields the overall free energy to insert q, part of it into the reactant state and part of it into the product state:

$$\Delta G(0\rightarrow q) = \lambda q \Delta G_s^{reac} + (1 - \lambda) q \Delta G_s^{prod} + \lambda^2 q^2 \Delta G_r^{reac} - (1 - \lambda)^2 q^2 \Delta G_r^{prod} \quad (6)$$

Generalization to a group of perturbing charges is straightforward. Different combinations (λ, $1 - \lambda$) give different weights to the static and relaxation terms at the two endpoints. The combination (1,0), for example, does not involve the product state contributions at all. Transformation from the reactant to the product state is made in a single step, using only the structure(s) associated with the reactant state. Structural changes in the product state relative to the reactant state (i.e., relaxation in response to q) are accounted for implicitly, through ΔG_r^{reac}. In contrast, for symmetrical half-steps ($\lambda = {}^1/_2 = 1 - \lambda$), the free energy change takes the form

$$\Delta G(0\rightarrow q) = q/2(\Delta G_s^{reac} + \Delta G_s^{prod}) + q^2/4(\Delta G_r^{reac} - \Delta G_r^{prod}) \quad (7)$$

An important point is that the reaction potentials V_1^r at the endpoints, and therefore ΔG_r^{reac},ΔG_r^{prod}, can differ, even though the magnitude $q_u = 1$ of the source charge is the same. The differences can be shown to arise from terms of third-order or more in the exact expression of the charging free energy,[33] i.e., from a breakdown of linear response. If q is not too large, conformational differences between reactant and product states are presumably moderate and linear response is presumably accurate; the second term in eq 7 is then small, and $\Delta G(0 \rightarrow q)$

(with symmetrical half-steps) depends mainly on the static free energies. In that case, relaxation in the product state is mainly accounted for through the first term in eq 7. The different values of ΔG_s^{reac} and ΔG_s^{prod} are obtained below from simulations of the reactant and product states, respectively, i.e., by simulating the structural relaxation explicitly. Several recent linear response studies of charge insertion in proteins have used eq 7 without the (presumably small) relaxation term.[14,25] If the continuum model were exact, the free energy $\Delta G(0 \rightarrow q)$ in eq 6 would not depend on the choice of λ. Requiring this independence yields the important relations

$$\Delta G_r^{reac} = \Delta G_r^{prod} = \frac{1}{2q}(\Delta G_s^{prod} - \Delta G_s^{reac}) \qquad (8)$$

Starting from the midpoint and perturbing to the product state in two fractional steps would lead to equivalent relations. These relations follow from the assumption of linear response, and are thus more general than the dielectric continuum assumption. They have been used in the context of electron transfer theory[34] and fluorescence.[35] Indeed, the relaxation free energy is proportional to the fluorescence Stokes shift, while the static free energy change is related to the energy difference for absorption or emission relative to absorption or emission in a vacuum (see, e.g., eq 7 of ref 35). Equation 8 states that the relaxation free energy in response to insertion of a charge is the same for the reactant, product, and all intermediate states (since q is arbitrary) and that it is proportional to the static free energy change. These relations connect a fully implicit description of the dielectric relaxation (involving ΔG_r^{reac} or ΔG_r^{prod}) to a partly explicit one (involving $\Delta G_s^{prod} - \Delta G_s^{reac}$), i.e., one where structural relaxation of the protein is included explicitly but solvent is still treated implicitly. Equation 8 is verified approximately in the application below if, and only if certain values for the dielectric constants are used, namely a protein/ligand dielectric constant of one for ΔG_s and 4−8 for ΔG_r. These values are shown below to be physically reasonable for the present system. This does not necessarily mean that they will be appropriate for other cases (see Discussion).

The step combination (1,0) (or (0,1)) inserts the full charge q (or $-q$) into the reactant (or product) state. If $|q|$ is large, e.g. $|q| = +e$ as below, the linear response assumed by the continuum model is known to break down, both for small solutes in water[30,36,37] and charges in proteins.[7,25] More precisely, the reaction potential in response to q may still be proportional to q over a certain range, but with a reduced slope compared to the small q case; this behavior was found in water simulations for $e < |q| < 3e$.[38] In that case, saturation occurs, so that a smaller dielectric constant would be appropriate. Similarly, if $\lambda \gg 1 - \lambda$, a smaller dielectric constant might be appropriate to describe relaxation in the first perturbation step than the second. In the present application, dielectric saturation effects do not appear to be significant, since the same dielectric constant works well for the entire range $0 \leq \lambda \leq 1$.

3. Methods

3.1. Charge Insertion in the Protein. The charge insertion mimics qualitatively an Asp → Asn transformation. The "mutation" is accomplished by inserting partial charges (totalling $+e$) at appropriate atomic positions on a fixed ligand in the active site. To be consistent with the earlier MDFE study,[21,22] we do not allow protein titratable groups to become ionized or neutral in response to the mutation; a similar approach was used in simulating the titration of Glu58 in Rnase T1.[39]

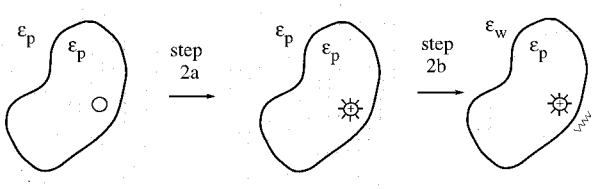

Figure 4. Two-step process for calculating the self-energy of a perturbing charge in a protein. Step 2a ("protein step") introduces the charge in an infinite protein medium; step 2b ("solvent step") introduces solvent around the protein. Permanent charges are shown as circles; induced volume and surface charge are schematized respectively as "prickles" around the permanent charges and as "squiggles" at the protein surface. The surface charge is schematically shown near the inserted charge, where it is largest. A counterion cloud can be present in the final state, but is not shown.

Pathways. The charge insertion is carried out in two main steps, schematized in Figure 2. A different protein dielectric constant is used for the two steps. In step 1 (static free energy), the potential on the ligand side chain is calculated with the protein charges and the (unperturbed) ligand charges as sources. A range of values are used for the protein dielectric constant ϵ_p^s ($\epsilon_p^s = 1-32$). Multiplying the potential at the sites of the perturbing charges by the perturbing charge values and summing over the sites gives the static free energy change ΔG_s.

The relaxation free energy step (step 2 in Figure 2) is itself decomposed into two steps, shown in Figure 4. In the first, step 2a, the perturbing charges are transferred from vacuum into a uniform "protein" medium with dielectric constant ϵ_p^r. In step 2b, the dielectric constant of the outer solvent region is changed from ϵ_p^r to $\epsilon_w = 80$. A range of values are used for ϵ_p^r ($\epsilon_p^r = 1-32$).

Charges. There is no rigorous way to mimic the Asp → Asn transformation by a simple insertion of partial charges, because the two side chains have different numbers of atoms. To determine whether the choice of perturbing charges plays a significant role in the present analysis, a range of perturbing charge distributions were studied. Results are presented here for two distributions, shown in Figure 5 along with the standard CHARMM22 Asp and Asn charge distributions.[17] The first (set 1) attempts to map the Asn side chain charge distribution onto the Asp side chain "as closely as possible", using the following ad hoc procedure. We consider the interactions of Asn and Asp with an Arg side chain, since in the AspRS active site, Asp and Asn are primarily recognized by a conserved arginine residue (Arg489). The potential of mean force between Asn and the Arg side chain is calculated with a continuum model, in a uniform medium having a dielectric constant of either 20 or 80. The same calculation is performed for Asp, and the Asp CG and OD2 charges are adjusted to achieve a potential of mean force similar to that of Asn (the other ligand charges being fixed at the values they have in Asn). The difference between this charge distribution and the original Asp distribution defines the perturbing charge set 1 (Figure 5). The second charge set (set 2) places a zero charge on OD2 (similar to the Asn ND2-H_2 group) and fixes the other charges at the values found in Asn. Several other reasonable charge sets were tried and gave similar results (not shown).

Structures. Finite-difference Poisson−Boltzmann calculations (FDPB) were done for multiple (usually ∼100) structures taken from molecular dynamics simulations of AspRS complexed with Asp, Asn, or the "Asn-like" ligands obtained by perturbing Asp with charge sets 1 and 2 (middle parts of Figure 5). The Asp:AspRS complex represents the reactant state for charge insertion using either charge set. The complexes with the

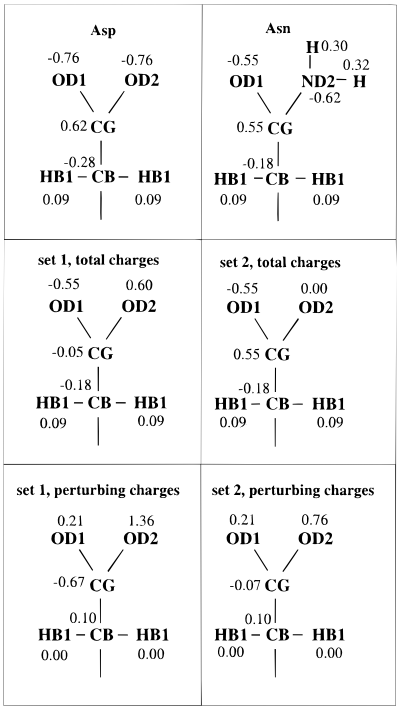

Figure 5. Upper row: CHARMM22 charge distributions of Asp and Asn before introducing the perturbing charges (units of *e*). Middle row: perturbed charge distributions, i.e., obtained by adding the perturbing charges from set 1 (left) or set 2 (right) to the Asp ligand. Bottom row: perturbing charge sets 1 and 2.

perturbed ligands represent the product states for the two charge insertions. Both product state structures are found to be very similar to the Asn:AspRS complex. FDPB results with either set of structures differ by a few percent at most. Therefore, the product structures and the Asn:AspRS structures can be viewed as essentially equivalent for our purposes. Most of the product state results reported below make use of the product structures. Calculations are also performed using thirty structures corresponding to a hybrid Asp/Asn ligand (the midpoint of the Asp → Asn transformation). The latter structures were taken from an appropriate 30 ps segment of an MDFE simulation.[22] In this simulation, the ligand has both an Asp and an Asn side chain, which typically occupy somewhat different positions. Below, both are used; i.e., the FDPB results are averaged over the two side chain positions for each of the 30 structures.

MD simulations of the Asp:AspRS and Asn:AspRS complexes in solution were described elsewhere.[21,22] These simulations were extended to give total durations of 600 and 500 ps, respectively. Simulations of the product states corresponding to charge sets 1 and 2 were carried out in the same conditions, and lasted 330 ps each. The latter simulations were initiated from the conformation reached after 100 ps of simulation of the Asn:AspRS complex; the Asn-like Asp was placed on top of the Asn ligand, which was removed. Both product state simulations were very stable and sampled conformations very similar to the Asn:AspRS complex. With charge set 1, for example, the ligand shifted at first by about 1 Å from its initial position (rms deviation averaged over all ligand atoms), then moved back to within 0.8 Å of the initial position and remained there for the final 200 ps. Surrounding side chains were equally stable.

Poisson−Boltzmann Calculations. The FDPB equation was solved for step 1 (static step) with a grid spacing of 0.40−0.60 Å, and the results shown to be robust with respect to grid spacing over this range. A grid spacing of 0.15 Å was used for step 2 (relaxation step), in which only a few permanent charges are present on the ligand (allowing a finer grid focusing). Atomic radii and charges were taken from ref 17. The protein−solvent boundary was defined by the molecular surface. A solvent probe radius of 2 Å was used. Most of the calculations were done in duplicate using the Delphi[40] and the UHBD[41] programs. Most calculations only included protein atoms contained within a spherical region of 20 Å radius centered on the ligand side chain, the rest of the protein being neglected; it is replaced by bulk solvent. A few calculations were done with the entire protein, demonstrating that this spatial cutoff introduced an error of only about 1 kcal/mol (∼ 1%; data not shown) in the charging free energy.

3.2. Choice of Dielectric Constant for the Ligand Side Chain. For the present charge insertion, a specific question arises concerning the initial state of the perturbing charges. The analysis above assumed the perturbing charges are initially in a vacuum. This places a constraint on the description of the protein in the FDPB calculations. Indeed, the latter require that the charges be embedded in the same dielectric medium in the starting and final states. If not, the results become dependent on charge self-energies that are ill-defined physically.[41] One way of satisfying this condition is to treat the ligand side chain as a small cavity in the protein, with a dielectric constant of one, so that it remains the same as in a vacuum. We refer to this as the "cavity" approach. Alternatively, one might view the ligand side chain as a small piece of the protein medium, with a dielectric constant ϵ_p^s or ϵ_p^r. In that case, the charges in the initial state must also be embedded in the same piece of dielectric medium. In either approach, the ligand backbone can be treated as a piece of the protein medium because its charges do not change.

Both approaches have drawbacks. The second approach is not optimal for the Asp → Asn transformation in solution (Simonson, T.; Archontis, G.; Karplus, M., unpublished results), leading to poor agreement with MDFE calculations. Therefore, it creates a difficulty when using a thermodynamic cycle such as Figure 1. In addition, with this approach, the free energy of the initial state becomes dependent on ϵ_p^s, and the overall results therefore become dependent on ϵ_p^s in an artificial way, through a state that does not involve the protein at all. Also, from a practical point of view, this approach complicates the static free energy FDPB calculation considerably; i.e., in the initial state there is induced charge at the surface of the ligand side chain, and this charge must also be included in the protein state (since rearrangement of induced charge is not allowed in the static step). The "cavity" approach is optimal for the solution transformation, but it has a practical drawback when used with a protein since it, too, requires a rather complicated FDPB protocol. Indeed, with this approach, three distinct dielectric media are present in steps 1 and 2b of the calculation: the ligand side chain cavity, the rest of the ligand + the protein, and the solvent. This is not feasible with the finite difference programs used here (Delphi[40] and UHBD),[41] which only allow two dielectric constants at a time. Therefore, for charge insertion in the protein, the "cavity" approach could be implemented only in an approximate way, as described in the Appendix. Numerical tests reported there for model systems with simple geometries suggest that the potentials obtained are accurate to within about 1−2%. This is of the same order of magnitude as variations that occur in the various calculational models below, e.g., when the solvent probe radius is changed from 1.5 to 2 Å, and does not affect the questions of interest. In the present application, the most consistent results (and the best

agreement with MDFE) are obtained using $\epsilon_p^s \approx 1$ and $\epsilon_p^r \approx 4-8$. With $\epsilon_p^s = 1$, the two approaches above are identical for the static step. For the relaxation step, the two approaches give similar results, especially in the low ϵ_p^r range (≤ 4). Therefore, in what follows, we report results with the "cavity" approach unless otherwise noted. This approach was also used in a study of charge insertion into cytochrome *c*.[7]

3.3. Molecular Dynamics Energies and Relaxation Free Energies. To compare with the static term in eq 3, the average interaction energy between the perturbing charges on the ligand side chain and the rest of the system was also estimated from molecular dynamics, including the effect of explicit solvent. This energy represents the linear response (MDLR) estimate of the static free energy:

$$\Delta G_s^{MD} = \langle U_{pert} \rangle_0 = \langle \sum_i q_i v_i \rangle_0 \qquad (9)$$

where the sum is over the perturbing charges q_i, v_i is the instantaneous electrostatic potential on q_i from the entire system except the ligand side chain, and the brackets indicate an average over an MD simulation of Asp:AspRS or Asn:AspRS. The MD simulations of refs 21 and 22 extended to 600 and 500 ps as described above, were used. Similarly, the linear response estimate of the relaxation free energy in response to the perturbing charges was calculated[7,13,42]

$$\Delta G_r^{MD} = -\frac{1}{2kT}(\langle U_{pert}^2 \rangle_0 - \langle U_{pert} \rangle_0^2) \qquad (10)$$

The calculations were performed for both the Asp:AspRS and the Asn:AspRS complexes.

The reactant → product transformation can also be divided into two-halfsteps, one from each endpoint. As in the continuum model, the static free energy above scales linearly with the step size (i.e., with the perturbing charge), while the relaxation free energy scales quadratically. The linear response approximation is expected to be better for each halfstep than for either full step; therefore, the MDLR free energy result with two-halfsteps should be more accurate than either full-step result. The MD model is a 20 Å sphere surrounded by vacuum,[21] whereas the continuum calculations include an infinite, surrounding, solvent continuum. To obtain the free energy for charge insertion in the presence of the infinite solvent continuum, we use the approach described in ref 21. The calculation involves three steps in all: first, the MD sphere is removed from bulk solvent to vacuum; second, the perturbing charges are inserted with the MDLR approach described above; third, the MD sphere is transferred back into bulk solvent. The free energies for the transfer steps were calculated earlier using continuum electrostatics (Figure 1 and Table 3 of ref 21, "protocol B"). Their total contribution (14 kcal/mol) is added to the MDLR charge insertion result for the MD sphere in a vacuum.

To characterize the polarizability of the active site region, following ref 7, the relaxation free energy was also calculated in response to a unit test charge placed successively on each α carbon within 10 Å of the ligand side chain. The calculation was based on the 600 ps simulation of the Asp:AspRS complex.

4. Results

4.1. Charge Insertion into the Protein: Static Free Energy. The static free energy (see eq 3) is

$$\Delta G_s = \sum_i q_i V_i \qquad (11)$$

TABLE 1: Static Free Energy[a]

structures[b]	perturbing charge set[c]	protein dielectric constant ϵ_p^s = 1	protein dielectric constant ϵ_p^s = 2
reactant	1	168.8 (12.1)	82.9 (5.9)
midpoint	1	93.6 (13.4)	44.6 (6.8)
product	1	8.5 (7.2)	2.3 (5.1)
reactant	2	158.1 (9.9)	77.6 (4.8)
midpoint	2	95.5 (8.8)	45.5 (4.5)
product	2	37.0 (6.1)	16.6 (3.1)
⟨reactant⟩[d]	1	165.2	80.8
Asp-X-ray[e]	1	164.8	80.9

[a] Static contribution ΔG_s (kcal/mol) to the free energy for charge insertion. CHARMM22 partial charges are used for all protein atoms. [b] Results averaged over simulations of AspRS complexed with Asp (116 conformations, one per picosecond; standard deviations in parentheses), with neutral Asp (63 and 106 conformations, 1 per 5 and 2 ps), and with an Asp/Asn hybrid (60 conformations, 1 per 0.5 ps). [c] Perturbing charge distributions in Figure 5. [d] Result for the average Asp:AspRS MD structure. [e] Result for the Asp:AspRS X-ray structure.

where the sum is over the perturbing charges q_i (shown in Figure 5) and V_i is the electrostatic potential at q_i, calculated in the absence of the perturbing charges. The value of V_i depends on the choice of protein dielectric constant ϵ_p^s associated with the unperturbed state. Results are reported in Table 1 for a protein dielectric constant equal to 1 and 2; larger values are not included because they lead to a free energy change that is too small, as we show below. Averages over the ensemble of reactant (Asp:AspRS) or product conformations obtained from the MD simulations are reported, as well as results for the midpoint of the Asp → Asn transformation, the average Asp:AspRS MD structure, and the Asp:AspRS crystal structure.

There is a significant dispersion of the results as a function of model parameters. For example, assuming $\epsilon_p^s = 1$ and averaging over the Asp:AspRS conformations, the static free energy varies from 158.1 kcal/mol (perturbing charge set 2) to 168.8 kcal/mol (perturbing charge set 1). Increasing ϵ_p^s to 2 reduces the free energies by almost exactly two; other choices of ϵ_p^s have a corresponding effect (not shown). The Asn:AspRS structures give results very close to the product state structures (within 2 kcal/mol; not shown). As shown in Methods (eq 7), if one performs the perturbation in two-half-steps, one from each endpoint, the relaxation free energy terms for each half-step cancel approximately. The cancelation condition is accurately satisfied in the present system (within 1 kcal/mol; see below). Therefore, the average of the endpoint static free energies provides an estimate of the overall free energy change. With $\epsilon_p^s = 1$, results are 88.6 and 97.4 kcal/mol for perturbing charge sets 1 and 2, compared to 95 kcal/mol for ΔG(Asp→Asn) from MDFE. The midpoint results are 93.6 kcal/mol (set 1) and 95.5 kcal/mol (set 2). In contrast, any other choice of ϵ_p^s gives unrealistic results; e.g., with $\epsilon_p^s = 2$, the free energy is 65 kcal/mol or less; with $\epsilon_p^s = 4$, it is 22 kcal/mol or less (not shown). All these results support the use of a low protein dielectric constant ($\epsilon_p^s \approx 1$) for calculating the equilibrium potential on the ligand.

While averaging over the static free energies from the two endpoints thus gives reasonable results, taking the static free energy for either endpoint alone does not, because the relaxation free energy must be included (see next section). This point was emphasized recently by Warshel and co-workers.[14,23,24]

The much smaller static free energy obtained with the product structures (8−37 kcal/mol) reflects the decreased polarity of the Asn side chain's surroundings. This is partly due to Asn shifting into a less polar position. The effect of the positional

TABLE 2: Relaxation Free Energy[a]

step[b]	perturbation charges[c]	structures[d]	protein dielectric constant ϵ_p^r = 1	2	4	8	16	32
step 2a	1		0.0	−47.0	−72.7	−86.4	−93.5	−97.2
	2		0.0	−33.0	−50.6	−59.8	−64.6	−67.0
step 2b	1	reactant	−20.5 (4.6)	−9.7 (2.2)	−4.5 (1.0)	−2.0 (0.4)	na	na
	1	product	−39.5 (7.8)	−18.7 (3.7)	−8.5 (1.7)	−3.6 (0.7)	na	na
	2	reactant	−17.1 (3.8)	−8.1 (1.8)	−3.7 (0.8)	−1.6 (0.3)	−0.6 (0.1)	−0.2 (0.0)
	2	product	−32.9 (4.1)	−15.6 (2.0)	−7.1 (0.9)	−2.8 (0.4)	−1.1 (0.1)	−0.3 (0.0)
total	1	reactant	−20.5	−56.7	−77.2	−88.4		
	1	product	−39.5	−65.7	−81.2	−90.0		
	2	reactant	−17.1	−41.1	−54.3	−61.4	−65.2	−67.2
	2	product	−32.9	−48.6	−57.7	−62.6	−65.7	−67.3

[a] Relaxation free energy ΔG_r (kcal/mol). Missing entries marked "na" were not calculated, since they correspond to a range of dielectric constants that gives poor consistency and poor agreement with MDFE (see text). (The results for these dielectric constants can actually be extrapolated approximately from the other data in the Table, since the relaxation free energy scales approximately as $1 - 1/\epsilon_p^r$.) [b] Steps in Figure 4: infinite protein step (step 2a), solvent introduction step (step 2b), and total (step 2a + step 2b). [c] Perturbing charge distribution (Figure 5). [d] Results averaged over reactant (Asp:AspRS) or product conformations, as indicated (standard deviation in parentheses).

TABLE 3: Relations between Static and Relaxation Free Energies[a]

perturbation charges[b]	$\epsilon_p^s = 1$ $\frac{(\Delta G_s^{reac} + \Delta G_s^{prod})}{2}$	$\epsilon_p^s = 1$ $\frac{\Delta G_r^{reac} - \Delta G_r^{prod}}{4}$	$\epsilon_p^r = 4$ $\frac{\Delta G_r^{reac} + \Delta G_r^{prod}}{2}$	$\epsilon_p^s = 1$ $\frac{\Delta G_s^{prod} - \Delta G_s^{reac}}{2}$	$\epsilon_p^r = 8$ $\frac{\Delta G_r^{reac} + \Delta G_r^{prod}}{2}$
1	88.6	1.0 ($\epsilon_p^r = 4$)	**−79.2**	**−80.2**	−89.2
2	97.4	0.3 ($\epsilon_p^r = 8$)	−56.0	**−60.6**	**−62.0**

[a] The linear response eq 8 predict that $^1/_2(\Delta G_s^{reac} + \Delta G_s^{prod})$ should be close to ΔG(Asp → Asn), that $^1/_2(\Delta G_r^{reac} - \Delta G_r^{prod})$ should be small, and that $^1/_2(\Delta G_s^{prod} - \Delta G_s^{reac})$ should be close to $^1/_2(\Delta G_r^{reac} + \Delta G_r^{prod})$. [b] Perturbing charge distribution (Figure 5).

shift can be illustrated by calculating the free energy using the product conformations, but placing the perturbing charges at the position occupied by the Asp side chain in the Asp:AspRS structures. With perturbing charge set 2, and excluding the contribution of the ligand itself, this gives 121.2 kcal/mol, compared to 38.0 kcal/mol with the Asn position, i.e., the Asn position is much less "polar". Another effect is the decreased reaction potential produced by Asn in its environment, compared to Asp. Indeed, the equilibrium potential in both the reactant and the product states includes a contribution from induced charge located around the permanent charges of the ligand (see discussion following eq 3). Since Asn is less polar than Asp, there is less induced charge and this contribution to the equilibrium potential is smaller (in absolute magnitude). This effect is illustrated by the contribution of the ligand backbone charges to ΔG_s: 26.5 kcal/mol at the Asn endpoint vs 51.3 kcal/mol at the Asp endpoint. At the Asp endpoint, the ligand backbone spends more time in close interaction with its own side chain; i.e., it is "polarized" more strongly than in the Asn case.

4.2. Charge Insertion into the Protein: Relaxation Free Energy. The self-energy of the perturbing charge, or relaxation free energy, is calculated in two steps (Figure 4). Results for each step are listed in Table 2 for different values of ϵ_p^r. The relaxation free energy is larger at the Asn endpoint, due to partial solvent exposure of the ligand side chain; the Asp side chain is completely buried in the Asp:AspRS complex. The "infinite protein" step 2a makes a large contribution to the relaxation free energy, which is dominant for $\epsilon_p^r \geq 2$. The total relaxation free energy varies from −20 kcal/mol (Asp endpoint, $\epsilon_p^r = 1$) to ca. −97 kcal/mol ($\epsilon_p^r = 32$), with approximately a $(1 - 1/\epsilon_p^r)$ dependence on ϵ_p^r, as expected for a self-energy. As with the static term, the Asn:AspRS structures for either charge set give results very similar to the product structures (not shown).

The consistency condition (eq 8) predicts that the relaxation free energy to insert a charge is the same at both endpoints, and that it is proportional to the static free energy change between endpoints. With perturbing charge set 1, these two predictions are approximately verified if, and only if the combination $\epsilon_p^s = 1$, $\epsilon_p^r = 4$ is used. For perturbing charge set 2, the combinations $\epsilon_p^s = 1$, $\epsilon_p^r = 4$ and $\epsilon_p^s = 1$, $\epsilon_p^r = 8$ are both reasonable, with the latter being slightly better (Table 3). Thus, for perturbing charge set 1, the static free energy change between the reactant and product states is $(\Delta G_s^{prod} - \Delta G_s^{reac})/2 - 80.2$ kcal/mol, compared to −79.2 kcal/mol for the relaxation free energy (average over the two endpoints; compare columns 4 and 5 in Table 3). For a perturbation to the midpoint (Asp → Asp/Asn), we have $\Delta G_s^{mid} - \Delta G_s^{reac} = -75.2$ kcal/mol, also close to the relaxation free energy, in agreement with eq 8. Perturbing charge set 2 with $\epsilon_p^s = 1$, $\epsilon_p^r = 8$ gives similar agreement. These results are consistent with the use of $\epsilon_p^s = 1$ in the previous paragraph and suggest that $\epsilon_p^r \approx 4-8$ is optimal to describe the relaxation.

4.3. Charge Insertion into the Protein: Total Free Energy. Table 4 summarizes data from Tables 1 and 2, listing the overall charging free energy $\Delta G(0 \rightarrow 1)$ for the most relevant choices of dielectric constants and for some representative combinations of steps from either endpoint and the midpoint.

We first consider the level of agreement between $\Delta G(0 \rightarrow 1)$ from FDPB and ΔG(Asp→Asn) from MDFE. As already mentioned, calculations with two symmetric half-steps are seen to give good agreement with MDFE when $\epsilon_p^s = 1$, while any larger value of ϵ_p^s gives very poor agreement. Similarly, when a full step from the reactant endpoint is used, only the choice $\epsilon_p^s = 1$, $\epsilon_p^r = 4$ (charge set 1) or $\epsilon_p^s = 1$, $\epsilon_p^r = 8$ (charge set 2) leads to agreement with the MDFE result. When a full step from the product (Asn) endpoint is used, these choices also give good agreement.

We next consider the consistency conditions, eq 8. Charge insertion at either endpoint gives results in close agreement with each other when $\epsilon_p^s = 1$ and $\epsilon_p^r = 4$ (charge set 1) or 4−8

TABLE 4: Total Free Energy in Protein Complex: Results for Selected Sets of Parameters and Pathways[a]

step combination[b]	perturbation charges[c]	ϵ_p^s	static term	ϵ_p^s	relaxation term	total ΔG
reac → prod	1	1	168.8	2	−56.7	112.1
reac → prod	1	1	168.8	4	−77.2	**91.6**
reac → prod	1	1	168.8	8	−88.4	80.4
reac → prod	2	1	158.1	4	−54.3	103.8
reac → prod	2	1	158.1	8	−61.4	**96.7**
prod → reac	1	1	8.5	2	65.7[d]	74.2
prod → reac	1	1	8.5	4	81.2	**89.7**
prod → reac	1	1	8.5	8	90.0	98.5
prod → reac	2	1	37.0	4	57.7	94.7
prod → reac	2	1	37.0	8	62.6	**99.6**
halfsteps	1	1	88.6	4	1.0	**89.6**
quartersteps	1	1	91.1	4	0.25	**91.4**
Asp → Asn, MDFE[e]						95.1 ± 2.8

[a] Free energies (kcal/mol) correspond to the reactant (react) → product (prod) direction. [b] Perturbing steps performed either from one endpoint (*e.g.*, reactant → product), or symmetrically from both endpoints ('halfsteps'), or from the endpoints and the midpoint ('quartersteps'), as schematized in Figure 3. [c] Perturbing charge set (Figure 5). Results with $\epsilon_p^s = 1$, $\epsilon_p^r = 4$ (set 1) or $\epsilon_p^s = 1$, $\epsilon_p^r = 8$ (set 2) are in bold. [d] The positive sign corresponds to the free energy contribution in the reactant → product direction, *i.e.*, the term $-\Delta G_r^{prod}$ that appears in eq 6. [e] Asp → Asn transformation by MDFE from ref 22.

(charge set 2); i.e., the result is approximately pathway-independent, in accord with eq 8. More generally, the overall $\Delta G(0 \rightarrow 1)$ can be calculated with nonsymmetric steps (λ, $1 - \lambda$) from either endpoint (Figure 6); if the continuum model is consistent, the result should not depend on λ. This is approximately verified with the above dielectric constants, as shown in Figure 6a for both perturbing charge sets. Similar behavior was found with all other charge sets examined (not shown). For a given perturbing charge set, results for all possible step combinations are within ±2 kcal/mol of each other. In contrast, combinations of dielectric constants where $\epsilon_p^s > 1$ and/or $\epsilon_p^r < 4$ exhibit a very strong λ-dependence. Calculations using additional steps from the midpoint lead to the same conclusions, as shown in Figure 6b. With charge set 1, all possible combinations of perturbing steps involving the midpoint give free energy changes within ±2 kcal/mol of each other, and the surface in Figure 6b is very flat. Charge set 2 (and all the other charge sets examined) lead to the same conclusions.

4.4. Static and Relaxation Free Energies from Molecular Dynamics and Linear Response. The mean perturbation energy (eq 9) and the linear response estimate of the relaxation free energy (eq 10) were calculated from molecular dynamics simulations of the reactant (Asp:AspRS) and product (Asn:AspRS) complexes. Results are given in Table 5. To obtain the overall free energy for charge insertion into the protein in bulk solvent, it is also necessary to transfer the MD model from vacuum to bulk solvent, as described in Methods. The corresponding free energy contribution was calculated with a continuum model in ref 21 and is reported in Table 5.

The molecular dynamics portions of the free energy (static, relaxation columns in Table 5) depend strongly on the perturbing charge set. Nevertheless, they are broadly consistent with the FDPB calculations. For example, the values of U_{pert} at either endpoint are within 11 kcal/mol of the ΔG_s values from FDPB with either charge set (except for the product state with charge set 2, where the deviation is larger). The molecular dynamics relaxation free energies ΔG_r^{MD} are also strongly dependent on the perturbing charge set and are about 15% larger on average than the FDPB result (ΔG_r). For example, with charge set 1, the molecular dynamics relaxation free energy (averaged over

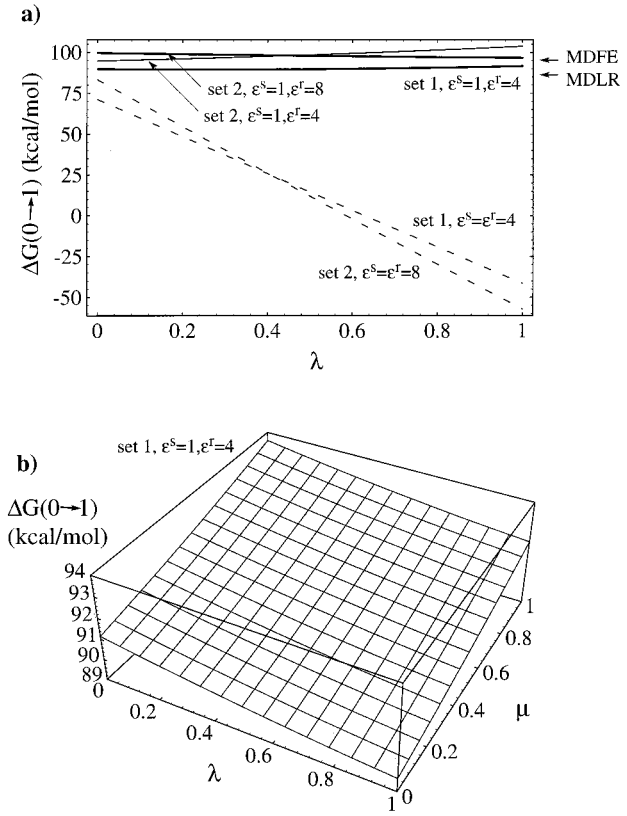

Figure 6. (a) Free energy $\Delta G(0 \rightarrow 1)$ for charge insertion in protein complex as a function of the step sizes λ, $1-\lambda$ from either endpoint (see text, eq 8). Results are shown for perturbing charge set 1, $\epsilon_p^s = 1$ or 4, $\epsilon_p^r = 4$, and for charge set 2, $\epsilon_p^s = 1$ or 8, $\epsilon_p^r = 4$ or 8. (b) Free energy $\Delta G(0 \rightarrow 1)$ for charge insertion in protein complex as a function of the step sizes from the Asp and Asn endpoints and the midpoint (see text, eq 8). Results are shown for charge set 1, $\epsilon_p^s = 1$, $\epsilon_p^r = 4$. A pair (λ, μ) in the plot corresponds to a pathway where the fractional charge $\lambda(q/2)$ is inserted into the reactant state, $(\lambda - 1)/q$ is removed from the midpoint state, $\mu(q/2)$ is inserted into the midpoint state, and $(\mu - 1)(q/2)$ is removed from the product state.

the two endpoints) is −97 ± 26 kcal/mol, compared to −79 ± 2 kcal/mol from FDPB (with $\epsilon_p^r = 4$). In addition, the MDLR relaxation corresponds to a 20 Å MD sphere; additional relaxation would occur in bulk solvent. A rough estimate of the additional relaxation is given by the Born free energy of a charge at the center of a 20 Å spherical cavity, ca. 8 kcal/mol; this is of the order of magnitude of the continuum contribution arising from the transfer of the MD model to and from bulk solvent (14 kcal/mol).

The total free energies from molecular dynamics are thus underestimated. However, if the MDLR free energy is calculated using a combination of two-halfsteps (one from each endpoint), so that the linear response approximation is expected to be more accurate, agreement with FDPB is much better: MDLR with halfsteps gives 84.7 and 85.9 kcal/mol for charge sets 1 and 2, compared to 90.0 and 97.7 kcal/mol from the FDPB calculations with halfsteps (with $\epsilon_p^s = 1$ and $\epsilon_p^r = 4$ or 8 as appropriate; see Table 3).

4.5. Polarizability of the Active Site from Molecular Dynamics and Linear Response. To further characterize the flexibility and polarizability of the active site, we apply a method proposed by Simonson et al.[7,12,13] A perturbing charge $q = 1$ is placed on a selected α carbon, and the MDLR relaxation free energy in response to this charge is calculated from eq 10. This provides a microscopic characterization of the active site polarizability beyond the ligand side chain, which can be

TABLE 5: Free Energies from Molecular Dynamics with Explicit Solvent[a]

structures[b]	transformation	perturbing charges[c]	static term[d]	relaxation term[e]	continuum correction[f]	total[g]
reactant	reac[j] → prod	1	172.0 (12.1)	−123.6	14.0	62.4
product	prod → reac	1	4.4 (9.2)	−71.2	−14.0	−80.8
reactant	reac → mid[h]	1	86.0	−30.9		
product	prod → mid[h]	1	2.2	−17.8		
	reac → mid → prod[i]	1	83.8	−13.1	14.0	**84.7**[i]
reactant	reac → prod	2	146.9 (9.5)	−76.8	14.0	84.1
product	prod → reac	2	0.2 (8.1)	−55.4	−14.0	−69.2
reactant	reac → mid[h]	2	73.4	−19.2		
product	prod → mid[h]	2	0.1	−17.8		
	reac → mid → prod[i]	2	73.3	−1.4	14.0	**85.9**[i]

[a] Molecular dynamics linear response (MDLR) free energies (kcal/mol). [b] Conformations used for averaging. [c] Perturbing charge set (Figure 5). [d] Static free energy $\langle U_{pert} \rangle$, where U_{pert} is the interaction energy between the perturbing charge distribution on the ligand side chain and everything else (protein + ligand backbone + explicit solvent); see eq 9. [e] Relaxation free energy $-1/2kT(\langle U_{pert}^2 \rangle - \langle U_{pert} \rangle^2)$. [f] Contribution of the transfer steps of the MD model to and from bulk solvent, calculated using continuum electrostatics (see Methods, ref 21). [g] Total free energy (static + relaxation + continuum) for introducing the transformation with MDLR. [h] Transformation to the midpoint. [i] The transformation is performed with two halfsteps. The free energy terms are thus obtained as the differences between the two-halfsteps immediately above. [j] reac = reactant, mid = midpoint, prod = product.

compared to other systems, such as those whose dielectric constant has been calculated from simulations. The ensemble average in eq 10 is over the unperturbed system (i.e., without the test charge); here, conformations are taken from a 600 ps molecular dynamics trajectory of the Asp:AspRS complex. Although the linear response approximation in eq 10 is not strictly valid for a perturbing charge as large as $q = 1$, the relaxation free energy scales as q^2, so the result can be written in kcal/mol/e^2 and viewed as a limiting result for small test charges. The calculation was done for the 38 C_α's within 10 Å of the ligand; the average relaxation free energy is -39.1 ± 9.8 kcal/mol/e^2. This includes the contributions of both protein and solvent. The contribution of transferring the MD model to and from bulk solvent must be added to this (see Methods). Rather than perform a precise calculation for each test charge position, we use a rough estimate given by the Born free energy of a perturbing charge in the 20 Å spherical cavity formed by the MD model, ca. 8 kcal/mol.[21] This gives a total of −47.1 kcal/mol. For comparison, the average over all C_α's in yeast cytochrome *c* is -68.0 ± 11.0 kcal/mol/e^2. The average over the C_α's in the inner half of cytochrome *c* is about half this value, or −34 kcal/mol/e^2. The AspRS value is intermediate. In cytochrome *c*, the protein dielectric constant that leads to the best fit between the relaxation free energies from an FDPB calculation and those from MDLR is 4 for all C_α's, and approximately 2 for the C_α's in the inner half.[7] Consistent with this, the Fröhlich−Kirkwood dielectric constant of cytochrome *c* from the same molecular dynamics trajectory is 2 for the inner half, and 4 for the bulk of the protein.[6,16] In summary, therefore, the average polarizability of the C_α positions in the AspRS active site is somewhat less than in cytochrome *c*, which is best characterized by a Fröhlich−Kirkwood dielectric constant of 4. This is slightly less than the range $\epsilon_p^r = 4-8$ that is optimal for charge insertion on the AspRS ligand. This may reflect the rigidity of the immediate environment of the C_α's, compared to the AspRS ligand, which can shift upon charge insertion.

5. Discussion

5.1. Macroscopic and Microscopic Descriptions of Dielectric Relaxation. The goal of this study is to analyze dielectric relaxation in an enzyme active site with the help of a continuum model. An important prerequisite is to determine a physically meaningful implementation of the continuum model, wherein the protein dielectric constant has a clear relation to the underlying structure and fluctuations of the system. To make this possible, we have focused on an especially simple process, the insertion of a small group of charges onto existing protein atoms. This process is simple enough so that static and relaxation free energies can be separated through a two-step pathway (Figure 2). At the same time, it mimics qualitatively the alchemical mutation of substrate Asp into Asn in the AspRS active site, studied earlier by molecular dynamics.[21,22] As mentioned above, it is difficult to carry out the exact Asp → Asn transformation with the continuum model, because this transformation requires the addition of two atoms in a position where they would overlap with a neighboring residue (Arg489); however, the charge insertion is sufficiently similar that the free energies can be compared in a meaningful fashion.

Different implementations of the continuum model were compared and tested in several ways. First, the continuum free energy change $\Delta G(0 \rightarrow 1)$ was compared with the alchemical free energy change ΔG(Asp→Asn) from MDFE. With the combination ($\epsilon_p^s = 1$, $\epsilon_p^r = 4-8$), agreement with MDFE is satisfactory, given that the transformations are not identical (charge set 1, 91 kcal/mol; charge set 2, 98 kcal/mol; MDFE, 95 kcal/mol); it is poor with almost all other choices of dielectric constants. The validity of $\epsilon_p^s = 1$ is in accord with the development of the molecular mechanics force field and charges. It indicates that these charges, combined with the MD snapshots of the reactant state structure, give a reasonable picture of the reactant charge distribution, without the need to include additional polarization of the protein implicitly. Certain other combinations of dielectric constants give similar results at the Asn endpoint, but never at both endpoints. At the Asn endpoint, ΔG_s and ΔG_r are both negative (in the Asn → Asp direction); increases in ϵ_p^s and ϵ_p^r change these terms in opposite directions and tend to cancel each other, so that several different combinations lead to similar free energies. At the Asp endpoint, however, ΔG_s is positive and ΔG_r negative; increasing ϵ_p^s beyond 1 reduces ΔG_s below ΔG(Asp→Asn), and relaxation with any value of ϵ_p^r can only decrease it further. This places much stricter restrictions on possible dielectric constants, and only a narrow range of combinations is possible at this endpoint.

The second set of comparisons tested the consistency conditions expressed by eq 8. The dielectric relaxation free energy calculated at either endpoint from the continuum model was compared to the static free energy change going to the other endpoint or the midpoint. In this way, fully implicit descriptions of the relaxation are compared to partly explicit ones. The combinations $\epsilon_p^s = 1$, $\epsilon_p^r = 4$ (charge set 1) or $\epsilon_p^s = 1$, $\epsilon_p^r = 8$ (charge set 2) give remarkably good agreement. This agreement,

by construction of eq 8, is a necessary and sufficient condition to achieve results that do not depend strongly on the starting point and choice of steps. "Standard" protocols that use a single dielectric constant $\epsilon_p^s = \epsilon_p^r \sim 2-8$ have an enormous path dependence (Figure 6). Agreement with MDFE for a particular path and protocol is likely to be fortuitous if eq 8 is not respected.

Third, the static and relaxation components of the charging free energy $\Delta G(0 \rightarrow 1)$ were calculated from MD with a linear response approximation (MDLR, eq 10). This calculation uses a second-order expansion of the charging free energy, which is only roughly valid for a perturbing charge of $+e$. Nevertheless, the MDLR free energies are broadly consistent with the FDPB results using $\epsilon_p^s = 1$, $\epsilon_p^r = 4$ or 8. The MDLR and FDPB results with halfsteps are particularly close.

Fourth, looking beyond the ligand side chain and probing the polarizability of the whole active site with a test charge/MDLR approach, the average polarizability is slightly smaller than that observed in the cytochrome *c* interior,[7] which has a Fröhlich−Kirkwood dielectric constant of about four. Given the less polar vicinity of the AspRS C_α's, compared to the vicinity of the AspRS ligand, this result is compatible with the optimal ϵ_p^r of 4−8 obtained above.

All of these comparisons indicate that for the simple charging process considered here, the continuum model can be applied in a way that is "nonempirical," i.e., the correct dielectric constants can be inferred from the underlying structures and fluctuations. This point had been demonstrated previously for small test charges ($|q| \leq e/4$) in cytochrome *c*.[6,7] The conclusions suggest an implementation of FDPB that is different from standard ones: the static and relaxation free energies are calculated using different values of the protein dielectric constant, as proposed by Krishtalik and others.[8,15] This result might appear physically inconsistent. In a fully consistent continuum model, one might expect a single protein dielectric constant to be sufficient to describe not only the present charge insertion process but also the assembly of the entire protein from its isolated atoms, or at least from small chemical groups, as well as the charge distribution of these groups in a vacuum. In fact, the need for two dielectric constants stems from the use of molecular mechanics charges in the FDPB context. Such charges, combined with the equilibrium structure, are well-suited to describe the equilibrium charge distribution in the protein, without the need for additional implicit protein polarization. Some polarization of the chemical groups within the protein is built into the charge set in an average way, since the charges were chosen to obtain the correct interactions for molecular dynamics simulations. The protein's response to a perturbing charge is poorly described by use of the unperturbed, equilibrium structure with the molecular mechanics charges. Instead, protein relaxation through displacement of the polar atoms has to be included. For the unperturbed structure, a dielectric constant greater than one is needed to describe this response. Because of the linearity of the Poisson−Boltzmann equation (at low ionic strength), the relaxation free energy is independent of the permanent charge distribution, so the different requirements for describing the equilibrium state and relaxation can be met with the two-step charge insertion of Figure 2.

5.2. Relevance to More Complicated Processes. Few biochemical processes are as simple as the charge insertion considered here. In a real ion-binding process, for example, desolvation of the binding site is required, and nearby ionizable groups can become charged or neutral in response to the binding. This makes it much more difficult to relate the dielectric constant(s) to the underlying structure and fluctuations. For example, systematic studies of charge−charge interactions were carried out recently,[24] which involved charge-pair creation, a process slightly more complicated than the present charge creation. That work showed that charge−charge interactions are screened by factors of 10 or more in proteins, relative to bare vacuum interactions. Most of the screening was shown to arise from the surrounding solvent, and no value of the protein dielectric constant for use in FDPB calculations of protein relaxation could be extracted from the data. In another study, pK_a shifts associated with the burial of lysines within the hydrophobic core of a protein were converted into dielectric constants through use of the Born equation.[43] The resulting dielectric constants depend not only on the polarizability of the hydrophobic core, but also on its polarity and on the proximity of solvent, and thus do not have a simple relation either to the dielectric constants introduced above (ϵ_p^r, ϵ_p^s), or to those used in more standard FDPB calculations.

The calculation of protein titration curves is even more complicated, involving simultaneous ionization of multiple groups. Thus, the use of a single protein dielectric constant $\epsilon_p \sim 20$ for these calculations is entirely empirical.[44] The approach developed above is not directly applicable when multiple sites can become ionized in response to insertion of the perturbing charge. However, an iterative approach might be appropriate. An initial calculation could be performed with a "standard" protocol (that uses $\epsilon_p \sim 20$ for the protein) to determine the ionization state of the titratable groups to a first approximation. A second calculation could then be performed to investigate the insertion or removal of a proton at a particular site of interest, using the protocol developed above and assuming the ionization states of the other titratable groups are fixed.

In Asp and Asn binding calculations to be reported elsewhere (vertical steps in Figure 1), a single protein dielectric constant of four is required to reproduce empirically the free energy ΔG(Asp→Asn) obtained from MDFE. This value is close to the average polarizability of the AspRS active site estimated above. Such a binding process can be decomposed into simpler steps, analogous to Figure 2: creation of a cavity, introduction of the ligand charges into the cavity with a fixed environment, and relaxation of the environment. The last two steps could presumably be modeled with protein dielectric constants of around 1 and 4, respectively. The cavity step is more complex: it would partly desolvate several charged groups in the AspRS active site and would depend both on the initial charge distribution and on rearrangement of induced charge around the new cavity. All these contributions behave differently as functions of ϵ_p; the optimal overall dielectric constant of four appears to be a compromise between these different effects. Calculations using one or more explicit cavity steps are needed to analyze the binding process in physical terms.

5.3. Polarizability of an Enzyme Active Site. Dielectric relaxation is a key aspect of charge transfer and catalysis in enzymes. Several authors have emphasized the need to treat it specifically in calculations of electrostatic free energies.[5,7,14,15,24] Different parameters and terminologies are found in the literature and so care must be used. In particular, it is important not to confuse polarity and polarizability, nor protein polarizability and solvent polarizability. Above, we used the dielectric constant ϵ_p^r as a consistent measure of protein polarizability.

The polarizability of the AspRS active site is moderate, corresponding to an ϵ_p^r of around 4−8. This range could not have been anticipated without performing the detailed analysis above. Knowledge of both the reactant and product state and

their fluctuations was critical. The resulting relaxation free energy is between −61 and −81 kcal/mol (depending mainly on the perturbing charge set used), including a contribution of between −2 and −8.5 from solvent. This is 15−35 kcal/mol less than the value that would be obtained in bulk water. For other charge rearrangement processes in the active site, such as those occurring during the enzymatic reactions, the relaxation free energies are also expected to be reduced compared to a solvent medium. If one assumes that the free energy curves for charge rearrangement in the corresponding reactant and product states are parabolic (as implied by linear response), reducing the relaxation (or reorganization) free energy is equivalent to a flattening of the two free energy parabolas.[10] If the overall reaction free energy is small (i.e., we are not in a Marcus inverted region), this always results in a lowering of their intersection. For a zero overall reaction free energy, for example, a reduction in relaxation free energy of 25 kcal/mol lowers the intersection by 25/4 ∼ 6 kcal/mol.[10] If the reaction is diabatic or weakly adiabatic (e.g., electron transfer), the activation free energy is lowered by the same amount. If the reaction is strongly adiabatic, the lowering of the activation free energy is expected to be smaller.

If the active site (protein and solvent) were viewed as a single, homogeneous, dielectric medium, we can ask what dielectric constant would be necessary to give the same relaxation free energy as the actual heterogeneous system. The relaxation free energy for the actual system is −79.9 kcal/mol (with charge set 1, $\epsilon_p^r = 4$, averaging over the two endpoints). For charge insertion into a homogeneous medium, relaxation free energy results are available in Table 2 (step 2a data). By interpolating the step 2a data in the table, we find that a uniform dielectric constant of about 5.2 is needed to reproduce the relaxation free energy (−79.9 kcal/mol) of the actual heterogeneous system.[52]

If the data for charge set 2 are used, a uniform dielectric constant of 8.7 is obtained. These dielectric constants are somewhat lower than the value of 10 assumed in a recent theoretical treatment of catalytic groups in serine proteases.[45] For comparison, the inner half of cytochrome *c*, from MD simulations, has a Fröhlich−Kirkwood dielectric constant of about 2;[6,16] as noted, this dielectric constant is directly comparable to ϵ_p^r (but not to ϵ_p^s). Such a low polarizability is required biologically to reduce the reorganization energy and consequently the activation free energy for electron transfer. A low dielectric relaxation (with obvious functional relevance) was observed experimentally in the photosynthetic reaction center in response to electron transfer, from Stark effect measurements before and after transfer;[46] the observed relaxation is compatible with a dielectric constant ϵ_p^s of 2−4. The active site of trypsin was found (from simulations) to have an average dielectric constant of about 8,[29] in good agreement with ϵ_p^r the present work.

6. Conclusions

The charging process considered here should be representative of proton binding and oxidation/reduction events in proteins, and qualitatively representative of simple alchemical changes such as Asp ↔ Asn or Glu ↔ Gln. The results suggest that, for such charging processes, the continuum model can be implemented in a "non-empirical" way, i.e., using two protein dielectric constants that have a clear relation to the underlying structure and fluctuations of the protein. They are based on a two-step charge insertion pathway commonly invoked in electron transfer theory. The first step, insertion into a fixed environment, yields the static free energy, which depends on the equilibrium reactant state potential. For the present system, this potential is given accurately by the molecular mechanics charge distribution, the equilibrium reactant state structure, and a protein dielectric constant of one, without the need to include additional protein polarization implicitly. The second step, relaxation of the environment, yields the relaxation free energy, which depends only on the shape and dielectric constant of the protein, but not on the charge distribution that existed prior to the charge insertion. For the present case, comparison with molecular dynamics free energy simulations and an internal consistency condition showed that a protein dielectric constant ϵ_p^r of 4−8 is appropriate for the relaxation step. For other systems, several general methods can be used to determine ϵ_p^r. When both the reactant and product states are known, the dielectric relaxation free energy can be calculated from the changes in electrostatic potential observed going from the reactant to the product state. This is done through eq 8 (see refs 23 and 47 for another example). The dielectric constant ϵ_p^r can then be chosen to reproduce this relaxation free energy. A second approach is to relate ϵ_p^r directly to the equilibrium fluctuations of the reactant state, calculating it from Fröhlich−Kirkwood theory.[6,15,16,28−31] This approach requires a single (long) simulation of the reactant state, fully solvated, and typically yields an average dielectric constant for the entire protein or a spherical region around its center. Alternatively, the dielectric constant can be adjusted to reproduce a few selected relaxation free energies calculated from MDLR, for example in response to test charges at selected locations.[7,12,13] All of these methods to determine ϵ_p^r are subject to error, but the present study suggests that, when taken together, they are robust enough to determine a meaningful dielectric constant.

For more complex processes such as the binding of a sizeable ligand to a protein, the two-step pathway of Figure 7 is insufficient. Desolvation of the binding site is required, introducing additional components in the free energy, such as cavity terms. The dielectric constant used in "standard" protocols is an empirical compromise between these components; a more sophisticated multistep analysis is required to clarify their exact magnitudes.

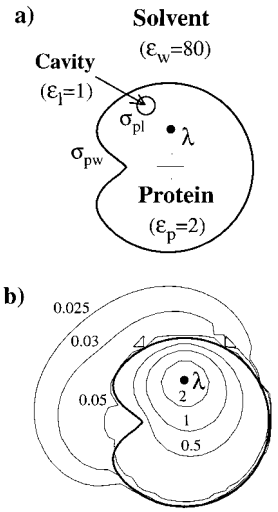

Figure 7. (a) Two-dimensional "protein" with an internal cavity (circle) and a charge λ (black dot), surrounded by solvent. The thick black line is the protein boundary, which has a Gaussian shape (see text). The origin of the coordinate system is shown by a cross. The *x* axis is horizontal; *y* is vertical. (b) Equipotential contours when no cavity is present, labeled by contour level (units of λ).

The protein polarizability in the AspRS active site was found here to be moderate, corresponding to a protein dielectric constant ϵ_p^r of around 4−8. Solvent in the active site increases the overall polarizability around the ligand only slightly; i.e., to obtain the same relaxation free energy in a homogeneous medium, a dielectric constant of about 5−8 would be required. This moderate polarizability is expected to lower the activation free energies for charge-transfer processes in the active site, compared to analogous processes in bulk water.[48] To understand the exact implications for catalysis, the detailed mechanism of AspRS must be studied.

Acknowledgment. We thank Michael Schaefer for helpful discussions. M.K. acknowledges partial support by the National Institutes of Health, the CNRS, and the Ministère de l'Education. T.S. acknowledges support from the CNRS. G.A. acknowledges support from the Cyprus Government through the grant "From Strong Interactions to Molecular Recognition: Theoretical and Computational Studies".

Appendix

The discussion of charge insertion above assumes that the perturbing charges are initially in a vacuum. Using an initial vacuum state places constraints on the FDPB calculations (Section 3.2); i.e., the most consistent FDPB implementation is one that treats the ligand side chain in the protein−ligand complex as a small cavity. This was done for part of the relaxation free energy calculation (step 2a, Figure 4); however, the effect of the cavity was neglected in the static free energy calculation and in step 2b of the relaxation free energy calculation (Figure 4). In this section, we analyze the error thus introduced, using two-dimensional model proteins. Calculations for such systems are simpler than for complex three-dimensional shapes.

We first consider calculations of the static free energy with and without a cavity surrounding the ligand side chain (Figure 7a). The reactant state electrostatic potential includes contributions from permanent charges on the ligand and in the protein, as well as from induced volume charge on the permanent charges and induced surface charge at all dielectric interfaces. Let ϵ_l be the ligand side chain dielectric constant and ϵ_p^s the dielectric constant of the ligand backbone and the protein, assumed to be greater than one. If $\epsilon_l = 1$ (cavity approach), the protein−ligand interface represents a dielectric discontinuity, with a certain distribution of induced surface charge $\sigma_{pl}(r)$ (where r represents the position within the interface). This surface charge gives rise to a certain potential $V_{pl}(r_i)$ at the perturbing charge positions r_i within the cavity. It also interacts with surface charge $\sigma_{pw}(r)$ at the protein−water interface (Figure 7a). If the cavity is sufficiently far from the permanent charges and the solvent, $\sigma_{pl}(r)$ will be small and the contribution of $V_{pl}(r_i)$ to the static free energy will be negligible. In addition, $\sigma_{pw}(r)$ will not be significantly influenced by $\sigma_{pl}(r)$, so that the contribution of $\sigma_{pw}(r)$ to the potential at the perturbing charge positions will be the same as with $\epsilon_l = \epsilon_p^s$. The magnitude of these effects was measured for two-dimensional model proteins in solution. The "two-dimensional" models considered are in fact extended along the third dimension to infinity; e.g., what appears in Figure 7a as a circular cavity is in fact an infinite cylinder perpendicular to the plane of the figure. The "point" charge λ in Figure 7a is in fact a line of charge perpendicular to the Figure, with density λ per unit length. The potential produced by such a line of charge in a uniform medium of dielectric constant ϵ_w is $-2\lambda/\epsilon_w \ln(r/r_0)$, where r is the distance from the line and r_0 is a constant defining the distance where the potential is zero. We consider here both a "circular" protein of radius 20 (in arbitrary units of length) and a "gaussian" protein (Figure 7) whose contour is defined in cylindrical coordinates by $r(\theta) = 20(1 - e^{-10(\theta-\pi)2})$. The constant r_0 is set to 100. Similar models were studied by Zauhar and Morgan.[49] The proteins contain one permanent charge located 5−6 length units from the protein−solvent boundary, either embedded directly in the protein, or surrounded by a small (2−3 length units) cavity. This geometry is roughly comparable to that of the ligand in the AspRS active site. The protein dielectric constant is taken to be two for both the static and relaxation steps. Potentials are expressed in units of charge per unit length, and the effect of the cavity on the potentials is measured relative to the total potentials. All calculations were done with the Mathematica program.[50]

TABLE 6: Effect of Cavity on Potential from an External Charge[a]

cavity position	r	$V(r)$ with cavity	$V(r)$ without cavity	difference
(0,15)	(0,14.5)	1.0658	1.1289	−6%
(0,15)	(0,15)	0.9436	0.9679	−3%
(0,15)	(0,15.5)	0.8310	0.8200	1%
(0,15)	(0.5,15.5)	0.9382	0.9606	−2%
(0,15)	(−0.5,15.5)	0.9394	0.9622	−2%
(−5,15)	(−5,14.5)	0.7175	0.6779	6%
(−5,15)	(−5,15)	0.6324	0.6128	3%
(−5,15)	(−5,15.5)	0.5469	0.5473	0%
(−5,15)	(−5.5,15.5)	0.5722	0.5669	1%
(−5,15)	(−4.5,15.5)	0.6931	0.6592	5%
(−10,15)	(−10,14.5)	0.2456	0.2473	−1%
(−10,15)	(−10,15)	0.2122	0.2155	−2%
(−10,15)	(−10,15.5)	0.1777	0.1820	−2%
(−10,15)	(−10.5,15.5)	0.1828	0.1848	1%
(−10,15)	(−9.5,15.5)	0.2427	0.2480	−2%

[a] Potentials (in units of charge per unit length) at the position r calculated for the "gaussian" protein shown in Figure 7. The source charge $\lambda = 1$ is located at (0,10). The cavity radius is 2, the protein dielectric constant is 2, and the solvent dielectric constant is 80. The potential is calculated with and without the cavity. The difference is expressed as a percentage of the result without the cavity.

Figure 7b shows the Gaussian "protein" with a single permanent charge located at (0,10) (the coordinate system is the one introduced above to define the protein boundary), along with equipotential contours calculated in the absence of the internal cavity. The potentials with and without a cavity are reported in Table 6 for three different cavity locations. The potentials with and without a cavity differ by less than 6% in all cases and by less than 3% in most cases. Calculations with a circular protein and other charge locations lead to similar results (not shown). These differences are small enough so that they do not affect the conclusions in the text concerning the poor consistency of the AspRS FDPB models when $\epsilon_p^s \geq 2$. For the most consistent models, where $\epsilon_p^s = 1$, there is no dielectric discontinuity at the protein−ligand interface, and no error is introduced for the static free energy by the cavity approach.

In the relaxation free energy calculation, the permanent charges are those on the ligand side chain. The approximate treatment of the ligand cavity is introduced in the second relaxation step (step 2b in Figure 4); i.e., the cavity is neglected in the FDPB calculations of step 2b above. If the cavity is sufficiently far from solvent, $\sigma_{pw}(r)$ (squiggles in Figure 4) does not significantly influence $\sigma_{pl}(r)$ (prickles in Figure 4), so that V_{pl} is the same in the initial and final states of step 2b. Conversely, $\sigma_{pl}(r)$ then does not affect $\sigma_{pw}(r)$, so that the reaction

TABLE 7: Effect of Cavity on Self-Potential from a Charge[a]

	with cavity[b]		infinite protein[c]	no cavity[d]
position r	$V_{pl}(r)$	$V_{pw}(r)$	$V_{pl}(r)$	$V_{pw}(r)$
(0,13)	−3.830 06	−2.139 39	−3.851 1 (0.021)	−2.164 99 (0.026)
(0,14)	−3.85 11	−2.201 87	−3.851 1 (0.000)	−2.228 97 (0.027)
(0,15)	−3.87 36	−2.268 71	−3.851 1 (0.022)	−2.297 51 (0.029)
(−1,14)	−3.850 37	−2.199 69	−3.851 1 (0.001)	−2.226 71 (0.027)
(1,14)	−3.850 37	−2.199 69	−3.851 1 (0.001)	−2.226 71 (0.027)
(0,13.5)	−3.840 41	−2.170 12	−3.851 1 (0.011)	−2.196 45 (0.026)
(0,14.5)	−3.862 16	−2.234 71	−3.851 1 (0.011)	−2.262 63 (0.028)
(−0.5,14)	−3.850 91	−2.201 32	−3.851 1 (0.000)	−2.228 41 (0.027)
(0.5,14)	−3.850 91	−2.201 32	−3.851 1 (0.000)	−2.228 41 (0.027)
(−0.5,14.5)	−3.861 96	−2.234 12	−3.851 1 (0.011)	−2.262 02 (0.028)
(−0.5,13.5)	−3.840 23	−2.169 61	−3.851 1 (0.011)	−2.195 92 (0.026)
(0.5,13.5)	−3.840 23	−2.169 61	−3.851 1 (0.011)	−2.195 92 (0.026)
(0.5,14.5)	−3.861 96	−2.234 12	−3.851 1 (0.011)	−2.262 02 (0.028)

[a] Reaction potential (in units of charge per unit length) at the position r calculated for a circular protein of radius 20 centered at (0,0). The cavity is centered at (0,14), the cavity radius is 2.2, and the source charge $\lambda = 1$ is at the center of the cavity. The protein dielectric constant is 2; the solvent dielectric constant is 80. [b] Potentials from induced charge located at the protein−cavity interface ($V_{pl}(r)$) and the protein−solvent interface ($V_{pw}(r)$). [c] Potential from induced charge located at the protein−cavity interface with an infinite protein (no solvent, middle state in Figure 4). In parentheses: unsigned difference with respect to column 2. [d] Potential from induced charge located at the protein−solvent interface with no cavity present. In parentheses: unsigned difference with respect to $V_{pw}(r)$ calculated with cavity (column 3).

potential from induced charge at the protein−solvent interface is not affected by the presence of the cavity. Therefore, the cavity does not affect the free energy change for this step and it can be neglected. On the other hand, if the solvent is closer, surface charge at the ligand−protein and protein−solvent boundaries will influence each other, the reaction potential will be modified, and the free energy will be affected by the presence of the cavity.

Results with a circular "protein" are given in Table 7. The potential produced within the cavity by the induced surface charge at the cavity−protein and protein−solvent interfaces is reported. The protein radius is 20, the cavity radius is 2.2 (corresponding to a carbon Born radius if the length units are in angstroms), and the cavity center is 6 length units away from solvent, similar to the perturbing charges in AspRS. The cavity is seen to modify the reaction potentials by at most 1−2%. Thus, neglecting the cavity in this case leads to a 1−2% error in the free energy for step 2b (which itself represents less than 10% of the total relaxation free energy in the AspRS application above).

In summary, neglecting the effect of the cavity introduces only a small error for the two-dimensional model "proteins", and is expected to have a similarly small effect in the AspRS calculations above.

References and Notes

(1) Rogers, N. The modelling of electrostatic interactions in the function of globular proteins. *Prog. Biophys. Mol. Biol.* **1986**, *48*, 37−66.

(2) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144.

(3) Antosiewicz, J.; McCammon, J.; Gilson, M. The determinants of pK_a's in proteins. *Biochemistry* **1996**, *35*, 7819−7833.

(4) Demchuk, E.; Wade, R. Improving the continuum dielectric approach to calculating pK_a's of ionizable groups in proteins. *J. Phys. Chem.* **1996**, *100*, 17373−17387.

(5) Warshel, A.; Aqvist, J. Electrostatic energy and macromolecular function. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 268−298.

(6) Simonson, T.; Perahia, D. Internal and interfacial dielectric properties of cytochrome *c* from molecular dynamics simulations in aqueous solution. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082−1086.

(7) 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.

(8) Krishtalik, L.; Kuznetsov, A.; Mertz, E. Electrostatics of proteins: description in terms of two dielectric constants simultaneously. *Proteins* **1997**, *28*, 174−182.

(9) Fröhlich, H. *Theory of Dielectrics*; Clarendon Press: Oxford, 1949.

(10) Marcus, R. Chemical and electro-chemical electron transfer theory. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155−196.

(11) Landau, L.; Lifschitz, E. *Statistical Mechanics*; Pergamon Press: New York, 1980.

(12) Simonson, T.; Perahia, D.; Bricogne, G. Intramolecular dielectric screening in proteins. *J. Mol. Biol.* **1991**, *218*, 859−886.

(13) Simonson, T.; Perahia, D.; Brünger, A. T. Microscopic theory of the dielectric properties of proteins. *Biophys. J.* **1991**, *59*, 670−90.

(14) 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.

(15) Simonson, T.; Brooks, C. L. Charge screening and the dielectric constant of proteins: insights from molecular dynamics. *J. Am. Chem. Soc.* **1996**, *118*, 8452−8458.

(16) Simonson, T. The dielectric constant of cytochrome *c* from simulations in a water droplet including all electrostatic interactions. *J. Am. Chem. Soc.* **1998**, *120*, 4875−4876.

(17) 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.

(18) Baker, W.; Yager, W. The relation of dielectric properties to structure of crystalline polymers. II. Linear polyamides. *J. Am. Chem. Soc.* **1942**, *64*, 2171−2177.

(19) Bone, S.; Pethig, R. Dielectric studies of the binding of water to lysozyme. *J. Mol. Biol.* **1982**, *157*, 571−575.

(20) Bone, S.; Pethig, R. Dielectric studies of protein hydration and hydration-induced flexibility. *J. Mol. Biol.* **1985**, *181*, 323−326.

(21) Simonson, T.; Archontis, G.; Karplus, M. Continuum treatment of long-range interactions in free energy calculations. Application to protein−ligand binding. *J. Phys. Chem. B* **1997**, *101*, 8349−8362.

(22) 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.

(23) 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.

(24) Sham, Y.; Muegge, I.; Warshel, A. The effect of protein relaxation on charge−charge interactions and dielectric constants of proteins. *Biophys. J.* **1998**, *74*, 1744−1753.

(25) Del Buono, G.; Figueirido, F.; Levy, R. Intrinsic pK_a's of ionizable residues in proteins: an explicit solvent calculation for lysozyme. *Proteins* **1994**, *20*, 85−97.

(26) Landau, L.; Lifschitz, E. *Electrodynamics of Continuous Media*. Pergamon Press: New York, 1980.

(27) Warshel, A.; Russell, S. Calculations of electrostatic effects in biological systems and in solutions. *Q. Rev. Biophys.* **1984**, *17*, 283−342.

(28) Gilson, M.; Honig, B. The dielectric constant of a folded protein. *Biopolymers* **1986**, *25*, 2097−2119.

(29) King, G.; Lee, F.; Warshel, A. Microscopic simulations of macroscopic dielectric constants of solvated proteins. *J. Chem. Phys.* **1991**, *95*, 4366−4377.

(30) Smith, P.; van Gunsteren, W. Predictions of free energy differences from a single simulation of the initial state. *J. Chem. Phys.* **1994**, *100*, 577−584.

(31) Löffler, G.; Schreiber, H.; Steinhauser, O. Calculation of the dielectric properties of a protein and its solvent: theory and a case study. *J. Mol. Biol.* **1997**, *270*, 520−534.

(32) Delepierre, M.; Dobson, C.; Karplus, M.; Poulsen, F.; States, D.; Wedin, R. Electrostatic effects and hydrogen exchange behavior in proteins. the pH dependence of exchange rates in lysozyme. *J. Mol. Biol.* **1987**, *197*, 111−130.

(33) Archontis, G.; Karplus, M. Cumulant expansion of the free energy: application to free energy derivatives and component analysis. *J. Chem. Phys.* **1996**, *105*, 11246−11260.

(34) Marcus, R. On the theory of shifts and broadening of electronic spectra of polar solutes in polar media. *J. Chem. Phys.* **1965**, *43*, 1261−1274.

(35) Muino, P.; Callis, P. Hybrid simulations of solvation effects on electronic spectra: indoles in water. *J. Chem. Phys.* **1994**, *100*, 4093−4109.

(36) Dejaegere, A.; Karplus, M. Analysis of coupling schemes in free energy simulations: a unified description of nonbonded contributions to solvation free energies. *J. Phys. Chem.* **1996**, *100*, 11148−11164.

(37) Hummer, G.; Pratt, L.; Garcia, A. Multistate Gaussian model for electrostatic solvation free energies. *J. Am. Chem. Soc.* **1997**, *119*, 8523−8527.

(38) Jayaram, B.; Fine, R.; Sharp, K.; Honig, B. Free energy calculations of ion hydration: an analysis of the Born model in terms of microscopic simulations. *J. Phys. Chem.* **1989**, *93*, 4320−4327.

(39) Mackerell, A.; Sommer, M.; Karplus, M. pH dependence of binding reactions from free energy simulations and macroscopic continuum electrostatics calculations: application to the 2′GMP/3′GMP binding to Ribonuclease T1 and implications for catalysis. *J. Mol. Biol.* **1995**, *247*, 774−807.

(40) Nicholls, A.; Honig, B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson−Boltzmann equation. *J. Comput. Chem.* **1991**, *12*, 435−445.

(41) Davis, M.; Madura, J.; Luty, B.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187.

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

(43) Garcia-Moreno, B.; Dwyer, J.; Gittis, A.; Lattman, E.; Spencer, D.; Stites, W. Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **1997**, *64*, 211−224.

(44) Antosiewicz, J.; McCammon, J.; Gilson, M. Prediction of pH dependent properties of proteins. *J. Mol. Biol.* **1994**, *238*, 415−436.

(45) Furuki, T.; Sakurai, M.; Inoue, Y. Multidielectric description of electrostatic environment surrounding a bound substrate in enzyme systems. *J. Phys. Chem.* **1995**, *99*, 12047−12053.

(46) Steffen, M.; Lao, K.; Boxer, S. Dielectric asymmetry in the photosynthetic reaction center. *Science* **1994**, *264*, 810−816.

(47) Churg, A.; Weiss, R.; Warshel, A.; Takano, T. On the action of cytochrome *c*: correlating geometry changes upon oxidation with activation energies of electron transfer. *J. Phys. Chem.* **1983**, *87*, 1683−1694.

(48) Krishtalik, L. Catalytic acceleration of reactions by enzymes. Effect of screening of a polar medium by a protein globule. *J. Theor. Biol.* **1980**, *86*, 757−777.

(49) Zauhar, R.; Morgan, R. A new method for computing the macromolecular electric potential. *J. Mol. Biol.* **1985**, *186*, 815.

(50) Wolfram, S. *Mathematica. A System for Doing Mathematics by Computer*; Addison-Wesley: New York, 1991.

(51) For a perturbing charge density ρ_f, the interaction free energy with the reaction field is

$$\Delta G_r = \frac{1}{2}\int \rho_f V_{reac}\, d^3r = \int \frac{1}{2}\int \rho_{ind} V_f\, d^3r = -\frac{1}{2}\int V_f \mathrm{div}\mathbf{P}\, d^3r$$

where the integration is over all space, the second equality uses a standard reciprocity relation, **P** is the polarization density, ρ_{ind} is the induced charge (concentrated near ρ_f and the solute−solvent interface), V_{reac} is the reaction potential, and V_f is the direct potential produced by ρ_f, not including the reaction potential. Using $\mathrm{div}(V_f\mathbf{P}) = V_f\mathrm{div}\mathbf{P} + \mathbf{P}\cdot\mathrm{grad}V_f$, the Green−Ostrogradski theorem, and the fact that V_f vanishes far from ρ_f, this becomes

$$\Delta G_r = -\frac{1}{2}\int \mathbf{E_f}\cdot\mathbf{P}\, d^3r = -\frac{1}{2}\int \mathbf{E_f}\cdot\chi(\mathbf{r})\mathbf{E_f}\, d^3r = -\frac{1}{2}\langle \mathbf{E_f}, \chi\mathbf{E_f}\rangle$$

where $\mathbf{E_f} = -\mathrm{grad}V_f$, $\chi(\mathbf{r}) = (\epsilon(\mathbf{r}) - 1)/(4\pi)$ is the real-space susceptibility, and the brackets on the right denote the usual product between (square-integrable) functions. This derivation assumes the solute−solvent interface has a finite width and the dielectric constant $\epsilon(\mathbf{r})$ is a smooth function of position; if desired, one can now make the interface width go to zero and transform the expressions accordingly, in which case the induced surface charge appears.

(52) To perform the interpolation, we assume the relaxation free energy for charge insertion into a homogeneous medium depends on the dielectric constant ϵ through a term $A(1 - 1/\epsilon)$, and the constant A is fitted to the step 2a data in Table 2.