

# Dielectric Relaxation in an Enzyme Active Site: Molecular Dynamics Simulations Interpreted with a Macroscopic Continuum Model

Georgios Archontis\*<sup>†</sup> and Thomas Simonson\*<sup>‡</sup>

Contribution from the Department of Physics, University of Cyprus, PO 20537, CY 1678 Nicosia, Cyprus, and Laboratoire de Biologie et Génomique Structurales (C.N.R.S.), Institut de Génétique et Biologie Moléculaire et Cellulaire, 1 rue Laurent Fries, 67404 Illkirch-Strasbourg, France

Received March 19, 2001. Revised Manuscript Received August 10, 2001

**Abstract:** Dielectric relaxation plays an important role in many chemical processes in proteins, including acid–base titration, ligand binding, and charge transfer reactions. Its complexity makes experimental characterization difficult, and so, theoretical approaches are valuable. The comparison of molecular dynamics free energy simulations with simpler models such as a dielectric continuum model is especially useful for obtaining qualitative insights. We have analyzed a charge insertion process that models deprotonation or mutation of an important side chain in the active site of the enzyme aspartyl-tRNA synthetase. Complexes with the substrate aspartate and the analogue asparagine were studied. The resulting dielectric relaxation was found to involve both ligand and side chain rearrangements in the active site and to account for a large part of the overall charging free energy. With the continuum model, charge insertion is performed along a two-step pathway: insertion into a static environment, followed by relaxation of the environment. These correspond to different physical processes and require different protein dielectric constants. A low value of  $\sim 1$  is needed for the static step, consistent with the parametrization of the molecular mechanics charge set used. A value of 3–6 (depending on the exact insertion site and the nature of the ligand) is needed to describe the dielectric relaxation step. This moderate value indicates that, for this system, the local protein polarizability in the active site is within at most a factor of 2 of that expected at nonspecific positions in a protein interior.

## 1. Introduction

Many chemical events in proteins involve charge separation or transfer: enzymatic reactions, photoexcitation of bound chromophores, electron transfer, proton binding and release, and binding of metal ions and other charged ligands.<sup>1–4</sup> In response to such an event, the protein and solvent relax, or reorganize, both through electronic polarizability and through the motions of charged and polar groups of atoms. This relaxation is a key component of protein energetics. A striking example is given by electron transfer in cytochrome *c* between ruthenated histidines, introduced artificially at the protein surface, and the central heme group:<sup>5</sup> the corresponding reorganization energy is about 1 eV, which is comparable to the redox potential difference spanned by the entire respiratory chain in cells. In general, an important feature of enzyme reactions is to limit undesired dielectric relaxation, especially for proton and electron-transfer steps. Indeed, the low polarizability of the protein interior compared to solvent helps to reduce activation barriers and increase catalytic rates<sup>6,7</sup> and is one of the most fundamental physical properties of proteins.

Computer simulations<sup>1,8–14</sup> have shown that dielectric relaxation in proteins is very complex, involving both protein and solvent degrees of freedom and including both fast, local, and slow, collective, motions. Several experimental studies have made use of spectroscopic probes<sup>15–20</sup> whose optical absorption and emission are sensitive to the dielectric properties of the environment. The  $pK_a$ 's of ionizable groups<sup>21</sup> and the changes in stability<sup>22</sup> or redox potential<sup>23,24</sup> upon mutating charged

(7) Warshel, A. *Computer Modelling of Chemical Reactions in Enzymes and Solutions*; John Wiley: New York, 1991.

(8) Churg, A.; Weiss, R.; Warshel, A.; Takano, T. *J. Phys. Chem.* **1983**, *87*, 1683–1694.

(9) Simonson, T.; Perahia, D. *J. Am. Chem. Soc.* **1995**, *117*, 7987–8000.

(10) Simonson, T.; Brooks, C. L. *J. Am. Chem. Soc.* **1996**, *118*, 8452–8458.

(11) Xu, D.; Phillips, C.; Schulten, K. *J. Phys. Chem.* **1996**, *100*, 12108–12121.

(12) Muegge, I.; Qi, P.; Wand, A. J.; Chu, Z.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 825–836.

(13) Basu, G.; Kitao, A.; Kuki, A.; Go, N. *J. Phys. Chem. B* **1998**, *102*, 2076–2084.

(14) Boresch, S.; Höchtel, P.; Steinhauser, O. *J. Phys. Chem. B* **2000**, *104*, 8743–8752.

(15) Pierce, D.; Boxer, S. *J. Phys. Chem.* **1992**, *96*, 5560–5566.

(16) Steffen, M.; Lao, K.; Boxer, S. *Science* **1994**, *264*, 810–816.

(17) Jordanides, X.; Lang, M.; Song, X.; Fleming, G. *J. Phys. Chem. B* **1999**, *103*, 7995–8005.

(18) Manas, E.; Wright, W.; Sharp, K.; Friedrich, J.; Vanderkooi, J. *J. Phys. Chem. B* **2000**, *104*, 6932–6941.

(19) Mertz, E.; Krishtalik, L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2081–2086.

(20) Xie, A.; Kelemen, L.; Hendriks, J.; White, B.; Hellingwerf, K.; Hoff, W. *Biochemistry* **2001**, *40*, 1510–1517.

(21) Dwyer, J.; Gittis, A.; Karp, D.; Lattman, E.; Spencer, D.; Stites, W.; Garcia-Moreno, B. *Biophys. J.* **2000**, *79*, 1610–1620.

(22) Schaefer, M.; Sommer, M.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 1663–1683.

\* Corresponding authors. E-mail: archonti@ucy.ac.cy and simonson@igbmc.u-strasbg.fr.

<sup>†</sup> University of Cyprus.

<sup>‡</sup> Laboratoire de Biologie et Génomique Structurales (C.N.R.S.), I.G.B.M.C.

(1) Warshel, A.; Russell, S. *Q. Rev. Biophys.* **1984**, *17*, 283–342.

(2) Cha, Y.; Murray, C.; Klinman, J. *Science* **1989**, *243*, 1325–1330.

(3) Moser, C.; Keske, J.; Warncke, K.; Farid, R.; Dutton, P. *Nature* **1992**, *355*, 796–802.

(4) Nandi, N.; Battacharyya, K.; Bagchi, B. *Chem. Rev.* **2000**, *100*, 2013–2046.

(5) Winkler, J.; Gray, H. *Chem. Rev.* **1992**, *92*, 369–379.

(6) Marcus, R. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155–196.

residues are also sensitive to both the polarity and polarizability of local protein regions. However, the experimental interpretation is rarely straightforward; for example, it is extremely difficult to separate the solvent and protein response without resorting to simulations.

To obtain a qualitative understanding, a powerful approach is to compare detailed, atomistic simulations with simpler, phenomenological models. In particular, one can determine the macroscopic continuum model that reproduces selected free energies given by molecular dynamics simulations. This approach was used to study dielectric relaxation in several proteins,<sup>9,25–27</sup> including relaxation in response to charge insertion in the active site of the enzyme aspartyl-tRNA synthetase.<sup>28</sup> It has become especially reliable with the introduction of accurate treatments of long-range electrostatics in molecular dynamics free energy simulations.<sup>12,29–32</sup> We apply it here to another charge insertion process in the same enzyme, modeling the deprotonation of an important active site residue. Charge insertion is performed along a two-step pathway introduced by Marcus:<sup>6,33</sup> insertion into a static environment, followed by relaxation of the environment. The earlier study<sup>28</sup> demonstrated that, to characterize dielectric relaxation in a meaningful way with a continuum model, the two steps must be treated separately, using different protein dielectric constants for each step. The dielectric constant for the relaxation step provides an unambiguous measure of the protein polarizability and of the magnitude of dielectric relaxation in the active site.

Aminoacyl-tRNA synthetases catalyze the first step in the translation of the genetic code, covalently attaching a specific amino acid to a cognate tRNA molecule.<sup>34</sup> For accurate translation, the synthetases must be highly specific for both the amino acid and the tRNA molecule. Engineering an aminoacyl-tRNA synthetase to modify its specificity and confer activity with respect to nonnatural amino acids has been proposed as a means to artificially expand the genetic code.<sup>35–37</sup> We showed earlier that aspartyl-tRNA synthetase (AspRS) binds its substrate aspartate (Asp) 15.3 kcal/mol more strongly than the neutral analogue asparagine (Asn),<sup>38</sup> mainly through interactions with four charged side chains in the active site. When one of these, Lys198, is either deprotonated or mutated to a neutral side chain, Asn binding is predicted to increase to almost the same level as Asp.<sup>39</sup>

(23) Varadarajan, R.; Zewert, T.; Gray, H.; Boxer, S. *Science* **1989**, *243*, 69–72.

(24) Gunner, M.; Alexov, E.; Torres, E.; Lipovaca, S. *J. Biol. Inorg. Chem.* **1997**, *2*, 126–134.

(25) Simonson, T.; Perahia, D.; Brünger, A. T. *Biophys. J.* **1991**, *59*, 670–90.

(26) King, G.; Lee, F.; Warshel, A. *J. Chem. Phys.* **1991**, *95*, 4366–4377.

(27) Sham, Y.; Muegge, I.; Warshel, A. *Biophys. J.* **1998**, *74*, 1744–1753.

(28) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1999**, *103*, 6142–6156.

(29) Hummer, G.; Pratt, L.; Garcia, A. *J. Am. Chem. Soc.* **1997**, *119*, 8523–8527.

(30) Darden, T.; Pearlman, D.; Pedersen, L. *J. Chem. Phys.* **1998**, *109*, 10921–10935.

(31) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 8349–8362.

(32) Simonson, T. *J. Phys. Chem. B* **2000**, *104*, 6509–6513.

(33) Marcus, R. *J. Chem. Phys.* **1956**, *24*, 979–989.

(34) Arnez, J.; Moras, D. *Trends Biochem. Sci.* **1997**, *22*, 211–216.

(35) Hohsaka, T.; Ashizuka, Y.; Sasaki, H.; Murakami, H.; Sisido, M. *J. Am. Chem. Soc.* **1999**, *121*, 12194–12195.

(36) Liu, D.; Schultz, P. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4780–4785.

(37) Wang, L.; Magliery, T.; Liu, D.; Schultz, P. *J. Am. Chem. Soc.* **2000**, *122*, 5010–5011.

(38) Archontis, G.; Simonson, T.; Moras, D.; Karplus, M. *J. Mol. Biol.* **1998**, *275*, 823–846.

In our previous study of dielectric relaxation in AspRS,<sup>28</sup> we considered the insertion of a positive charge onto the side chain of the Asp ligand, converting it into a neutral, Asn-like analogue. A continuum model with a protein dielectric constant of 5–6 was necessary to reproduce the dielectric relaxation observed in molecular dynamics free energy (MDFE) simulations. Here, we consider insertion of a negative charge onto the Lys198 side chain, mimicking either its deprotonation or its mutation to a neutral side chain such as leucine. Calculations are performed for both the AspRS:Asp and the AspRS:Asn complexes. They confirm that different continuum models are needed for the first, static, and the second, relaxation, steps of charge insertion and that the relaxation step reports a protein dielectric constant (for this somewhat different insertion site) of 3–6, reasonably close to the result for the previous (ligand side chain) site. This and the previous study represent arguably the first consistent characterization of dielectric relaxation in an enzyme active site with a continuum model. They indicate that the local protein polarizability in the AspRS active site is not much greater than that expected at nonspecific positions in a protein interior (i.e., within a factor of 2), as estimated, for example, from molecular dynamics simulations.<sup>40,41</sup>

The approach presented here should be applicable to any process involving charge binding or transfer in proteins and can be used to compare the polarity and polarizability of active sites in different classes of proteins, which may differ widely. For example, preliminary results for the electron-transfer protein cytochrome *c* suggest (consistent with earlier work<sup>8,9,12,42</sup>) that its active site is even less polarizable than the AspRS active site considered here (TS, unpublished data).

## 2. Methods

**2.1 Molecular Dynamics Free Energy (MDFE) Simulations.** The “mutation” of interest here inserts partial charges totalling  $-e$  onto the Lys198 side chain. The native and modified proteins are denoted K198 and K198n (n for “neutral”) in what follows.

The free energy of charge insertion is first calculated by MDFE simulations, following a procedure that takes into account long-range electrostatic interactions through a continuum model. According to the method developed and described in detail in ref 31, the mutation is performed in a two-step thermodynamic pathway. In the first step, the mutation is introduced gradually in a finite simulation system including a spherical region of the protein and several hundred explicit water molecules. Selected side chain charges are reduced to mimic bulk solvent screening and reduce structural perturbations that could result from the finite size of the system. In the second step, the scaled charges are restored to their original values, and the finite system is transferred to a bulk medium. The free energy for the second step is obtained from Poisson–Boltzmann calculations with a continuum model. The detailed molecular dynamics setup for the first step is described in subsection 2.5.

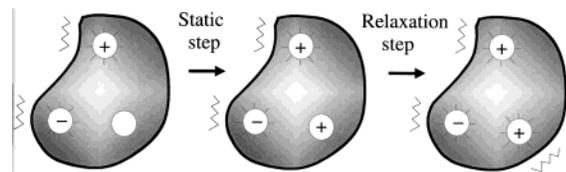
The subset of side chains to be scaled in the first step, and the exact scaling factors, can be chosen in a flexible manner, because the main objective is to reduce structural distortions in the simulations of the finite system. An earlier application of the method for the same protein demonstrated that it is robust with respect to the scaling protocol and the choice of scaled residues.<sup>31</sup> Here, side chains near the boundary of the simulation system not involved in salt bridges were chosen (this protocol is referred to as protocol B in ref 31). The scaling factors were obtained as the ratio of the electrostatic potentials on the K198 side chain when the AspRS:Asp complex (including the simulation water) is in vacuum or in solution.<sup>31</sup>

(39) Archontis, G.; Simonson, T.; Karplus, M. *J. Mol. Biol.* **2001**, *306*, 307–327.

(40) Simonson, T. *Int. J. Quantum Chem.* **1999**, *73*, 45–57.

(41) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243–252.

(42) Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368–8372.



**Figure 1.** Two-step pathway for introducing a perturbing charge into a protein. The protein and solvent are treated as two distinct dielectric media. The first, static step introduces the charge into a fixed environment; the second step allows the environment to relax. In the model used here, permanent charges (represented by + or -) do not move; relaxation is described by a rearrangement of induced polarization charge (represented by pricksles around the permanent charge and squiggles on the protein surface).

During the MD simulations (the first step of the present thermodynamic pathway), the charges required to convert Lys198 to its neutral analogue (or vice versa) were inserted gradually in increments of  $0.1e$  or  $0.05e$ . After each increment, the system was equilibrated for 40–60 ps, and then, data were collected for 40–60 ps, giving a total of 1000–2000 ps per mutation run. The charge insertion was performed for both the Asp:AspRS and Asn:AspRS complexes. The starting structures for the simulations were taken from various points of long (several hundred picoseconds) equilibrium simulations of the K198 and K198n states.

**2.2 Charge Insertion with a Continuum Model: Static and Relaxation Steps.** Charge insertion is also performed with a continuum model. In this context, the insertion is decomposed into two steps, following a procedure introduced by Marcus<sup>6,33</sup> and applied recently to the present system.<sup>28</sup> The two steps are depicted in Figure 1. In the first step, the new charges are inserted with all preexisting permanent and induced charges held fixed. The corresponding static free energy is

$$\Delta G_s = \sum_i q_i V_i \quad (1)$$

where  $q_i$  are the inserted charges and  $V_i$  is the electrostatic potential at the insertion site  $i$  in the absence of the new charges  $q_i$ . The protein dielectric constant has a value  $\epsilon_p^s$ , and the solvent has a dielectric constant  $\epsilon_w$ . A reference free energy has been subtracted, corresponding to charge insertion on an isolated Lys side chain in vacuum. This reference state is consistent with the procedure used in the MDFE calculations (see ref 28 for a discussion).

In the second step, the system is allowed to relax in response to the new charges. All permanent charges are still held fixed, so that the relaxation is entirely modeled by a redistribution of polarization charge. This is a common assumption for continuum models applied to proteins.<sup>43,44</sup> The corresponding relaxation free energy  $\Delta G_r$  is identical to the Born self-energy of the inserted charges.<sup>9</sup> For technical reasons, the relaxation free energy is itself calculated in two stages, as explained in refs 28 and 45 (see, e.g., Figure 4 of ref 28). In the first stage, the charges are transferred from vacuum into a cavity of dielectric constant unity, corresponding to the Lys198 side chain, embedded in an infinite “protein” medium with a dielectric constant of  $\epsilon_p^r$ . In the second stage, the dielectric constant of the outer solvent region is changed from  $\epsilon_p^r$  to  $\epsilon_w = 80$ .

It should be noted that this charging procedure is not exactly the original one of Marcus;<sup>33</sup> there, the electronic polarization was allowed to rearrange in the first step. Here, all the polarization rearranges in the second step. As a result, the present relaxation free energy is related, but not identical, to the “outer-shell”, or dielectric-continuum part of the usual Marcus reorganization energy.<sup>46,47</sup> The present procedure leads

(43) Sharp, K.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *19*, 301–332.

(44) Schaefer, M.; Vlijmen, H. V.; Karplus, M. *Adv. Protein Chem.* **1998**, *51*, 1–57.

(45) Hoefinger, S.; Simonson, T. *J. Comput. Chem.* **2001**, *22*, 290–305.

(46) Marcus, R. In *Protein Electron Transfer*; Bendall, D., Ed.; BIOS Scientific Publishers: Oxford, 1996, pp 249–272.

(47) Sharp, K. *Biophys. J.* **1998**, *73*, 1241–1250.

to a more straightforward parametrization of the continuum model when a molecular mechanics charge set is used. It is also closely related to the approach of Muegge et al.:<sup>12</sup> the present relaxation free energy is identical to their reorganization energy (e.g., Figure 1 of ref 12).

The protein dielectric constant  $\epsilon_p^r$  for the relaxation step provides a measure of the protein polarizability around the charge insertion site. As such, it should be comparable to the dielectric constant estimated from the protein dipole fluctuations in molecular dynamics simulations, which also measures the protein polarizability<sup>48,26,40</sup> (though the two are not necessarily identical, because the latter approach involves averaging the fluctuations over a large protein volume; see, e.g., ref 41 for a discussion).

**2.3 Charge Insertion with a Continuum Model: Compound Insertion Pathways.** To connect the reactant (K198) and product (K198n) states, one can insert the new charge  $q = -e$  into the K198 structure (in two steps, static + relaxation, as described previously); equivalently, one could remove it from the K198n structure. More generally, an infinite number of “compound” pathways are possible, as explained in ref 28. For example, we can insert  $\pm 1/2e$  charges into the K198n and K198 structures, respectively (pathway K198 → K198/K198n ← K198n, where K198/K198n denotes the midpoint state). In a general compound pathway, a fractional charge  $\lambda q$  is inserted into the initial reactant state, and  $(\lambda - 1)q$  is inserted into the product state; the total free energy change can be expressed within the framework of linear response as<sup>28</sup>

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

Here,  $\Delta G_s^{\text{react(prod)}}$  and  $\Delta G_r^{\text{react(prod)}}$  are the static and relaxation free energies to introduce a unit charge  $q_u = 1$  into the reactant (product) state.

Equation 2 can be used to express the free energy in terms of static and relaxation contributions along particular compound pathways. For two pathways involving half-charges, for example, and with  $q = -1$ , one gets

K198 → K198/K198n ← K198n:

$$\Delta G = -\frac{1}{2}(\Delta G_s^{\text{K198}} + \Delta G_s^{\text{K198n}}) + \frac{1}{4}(\Delta G_r^{\text{K198}} - \Delta G_r^{\text{K198n}}) \quad (3)$$

K198 ← K198/K198n → K198n:  $\Delta G = -\Delta G_s^{\text{K198/K198n}}$  (4)

$\Delta G_{s(r)}^x$  is the static (relaxation) free energy to insert a unit charge into state  $x$ . Equation 3 follows from eq 2 by taking K198 and K198n as reactant and product states, with  $\lambda = 0.5$ ; eq 4 is obtained by summing the free energies to add and to remove  $0.5q$  from the midpoint state K198/K198n.

More generally, if linear response is exact, the free energy  $\Delta G$  should not depend on the fraction of charge introduced into the reactant or product state, that is, on the parameter  $\lambda$  that appears in eq 2. This requirement yields the following important equations, which connect the static and relaxation free energies for a reactant and product state that differ by the charge  $q$ :<sup>28</sup>

$$\Delta G_r^{\text{react}} = \Delta G_r^{\text{prod}} = \frac{1}{2q}(\Delta G_s^{\text{prod}} - \Delta G_s^{\text{react}}) \quad (5)$$

These equations were first derived by Marcus in a somewhat different context.<sup>49</sup> They connect explicit and implicit descriptions of the dielectric relaxation and can be viewed as a fluctuation–dissipation theorem.<sup>50</sup> Because the static and relaxation free energy terms have a strong dependence on the protein dielectric constants  $\epsilon_p^s$  and  $\epsilon_p^r$ , eq 5 will usually be satisfied for only a few choices of  $\epsilon_p^s$  and  $\epsilon_p^r$ . The

(48) Gilson, M.; Honig, B. *Biopolymers* **1986**, *25*, 2097–2119.

(49) Marcus, R. *J. Chem. Phys.* **1965**, *43*, 1261–1274.

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

dielectric constants that best satisfy eq 5 represent the optimum parametrization of the continuum model (within the framework of linear response).

If corrections to the linear response approximation are thought to be important, the righthand side of eq 2 can be expressed in terms of the second- and higher-order cumulants of the charging energy,<sup>51</sup> with the higher-order cumulants giving rise to dielectric saturation effects. Dielectric saturation is neglected here, as in most applications of the continuum model to proteins. Nonlinear effects associated with counterions are also ignored.

Before proceeding, it should be noted that to exploit eqs 3–5 in a meaningful way, it is necessary, for each charge state, to use equilibrium structures that are representative of that state.<sup>12,39,52,53</sup> For example, for the K198  $\leftarrow$  K198/K198n  $\rightarrow$  K198n pathway, a structure of the midpoint state is needed. In this work, structures of the necessary states are obtained by MD simulations (see below).

It is useful to compare the present approach to earlier calculations of charging free energies with a linear response approximation. For example, the “half-charge” pathway embodied in eq 3 corresponds to a protocol used by Muegge et al.,<sup>12</sup> embodied in their eq 6 (which assumes, in addition, that the relaxation free energies in the reactant and product states are equal). Because this pathway averages over the reactant and product state structures, the calculated free energy depends mainly on  $\Delta G_{\text{r}}^{\text{K198}}$  and  $\Delta G_{\text{r}}^{\text{K198n}}$  and very weakly on the relaxation free energies  $\Delta G_{\text{r}}^{\text{K198}}$  and  $\Delta G_{\text{r}}^{\text{K198n}}$ . Therefore, the choice of the relaxation dielectric constant  $\epsilon_{\text{p}}^{\text{r}}$  is not important. In contrast, if one follows a pathway where the entire perturbing charge is inserted into the reactant state structure, the result will depend strongly on both  $\epsilon_{\text{p}}^{\text{s}}$  and  $\epsilon_{\text{p}}^{\text{r}}$ . The possibility to calculate the same free energy along different simple and compound pathways allows us to focus on either the static or the relaxation component of the charging free energy, or both.

**2.4 Poisson Calculations.** To analyze charge insertion with the continuum model, finite-difference Poisson (FDP) calculations were performed for a large number of structures (100–200) for each charge state of residue Lys198. The structures were taken from MD simulations of each state (below). All atomic charges and radii were taken from the CHARMM22 parameter set,<sup>54</sup> with the exception of the hydrogen radii which were set to 1.0 Å. The FDP equation was solved with a cubic grid spacing of 0.4 Å. The molecular surface used to define the protein–solvent boundary was constructed using a probe sphere with a radius of 2 Å. With this probe radius, the protein has no internal cavities. The calculations were done with the UHBD program.<sup>55</sup>

**2.5 Molecular Dynamics Setup.** Equilibrium structures of the native (K198) and mutant (K198n) AspRS:Asp and AspRS:Asn complexes were obtained by MD simulations, using the CHARMM program.<sup>56</sup> The setup was described in detail elsewhere.<sup>38</sup> It consisted of an approximately spherical, 20 Å radius model which contained most of the active site of AspRS, the ligand, and 384 water molecules. The water was confined by a stochastic boundary potential. Electrostatic interactions were treated without truncation by use of a multipole expansion for groups more than 14 Å apart. The protein and water atoms in a buffer region (starting 15 Å away from the center of the sphere) obeyed Langevin dynamics and were subjected to random and frictional forces that mimic a thermal bath at 293 K.

In the mutant (K198n) complexes, the terminal CE and NZ charges of Lys198 were set to  $-0.18e$  and  $-0.27e$ , respectively, and the charges

**Table 1.** MDFE Results for K198  $\rightarrow$  K198n Mutation in AspRS:Asp and AspRS:Asn Complexes<sup>a</sup>

ligand	run	direction	alchemical MD step	continuum step <sup>b</sup>	total
Asp	1	forward	123.9	−29.6	94.3
Asp	2	forward	124.1	−29.6	94.5
Asp	3	backward	111.4	−29.6	81.8
Asp	4	backward	112.2	−29.6	82.6
Asp	5	forward	113.0	−29.6	83.4
Asp		average, forward runs 1–2 only			94.4
Asn	1	forward	90.5	−11.2	79.3
Asn	2	forward	96.8	−11.2	85.6
Asn	3	backward	84.4	−11.2	73.2
Asn	4	backward	83.1	−11.2	71.9
Asn	5	forward	87.4	−11.2	76.2
Asn		average, forward runs 1–2 only			82.5
Asn		average, all runs			77.2

<sup>a</sup> Free energies in kcal/mol. <sup>b</sup> Contribution corresponding to the continuum treatment of long-range interactions (see text).

of the hydrogen atoms connected to CE or NZ were set to 0.09e. This choice reproduces the atomic charges present on the Leu side chain in the CHARMM22 parameter set.<sup>54</sup> For the native AspRS:Asn complex structures, the simulations of ref 38 were used. For the K198n AspRS:Asn complex, additional simulations were performed.<sup>39</sup>

### 3. Results

We first present the results from MDFE simulations for charge insertion in the AspRS:Asp complex, and then, we determine the continuum model that is needed to reproduce them. MDFE and continuum results for charge insertion in the AspRS:Asn complex are presented next.

**3.1 Charge Insertion in the AspRS:Asp Complex: MDFE Simulations.** The free energy of the K198  $\rightarrow$  K198n transformation was first evaluated by MD free energy (MDFE) simulations. Results are listed in Table 1. Five runs were performed, three in the forward (K198  $\rightarrow$  K198n) and two in the backward (K198  $\leftarrow$  K198n) directions. The initial coordinates and velocities for forward runs 1 and 2 were taken from different points of a 500 ps trajectory of the K198 state. Backward runs 3 and 4 were started after, respectively, 200 and 400 ps of simulation at the end of forward run 1. Run 5 is discussed below.

The resulting free energy change is  $\Delta G_{\text{Asp}}(\text{K198} \rightarrow \text{K198n}) = 94.4 \pm 0.1$  kcal/mol from forward runs 1–2 and  $82.2 \pm 0.4$  kcal/mol from backward runs 3–4. While forward runs 1 and 2 converge to the same K198n structure (shown in Figure 2a), the backward runs do not return to the native K198 structure, but to a higher free energy structure, referred to as K198<sup>‡</sup> (shown in Figure 2b). Starting from this structure, another forward run was performed, run 5, which returned to essentially the same K198n structure as runs 1–2. Run 5 yielded a free energy difference in close agreement with backward runs 3 and 4 (Table 1). Thus, K198<sup>‡</sup> appears to be a high energy, metastable state that is reached preferentially and reversibly from K198n. A more elaborate protocol using biasing restraints would presumably be needed to drive the system reversibly from K198n to the low energy K198 state.<sup>38</sup>

The structural origin of the hysteresis between forward runs 1 and 2 and backward runs 3 and 4 is clear from an analysis of the residue free energy components along the various runs. The most important contributions are listed in Table 2. The large free energy values of runs 1 and 2 originate mainly from residues Asp233 and Gln231 and the ligand Asp. In the beginning of forward runs 1 and 2 (K198 state), Asp233 interacts with

(51) Archontis, G.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 11246–11260.

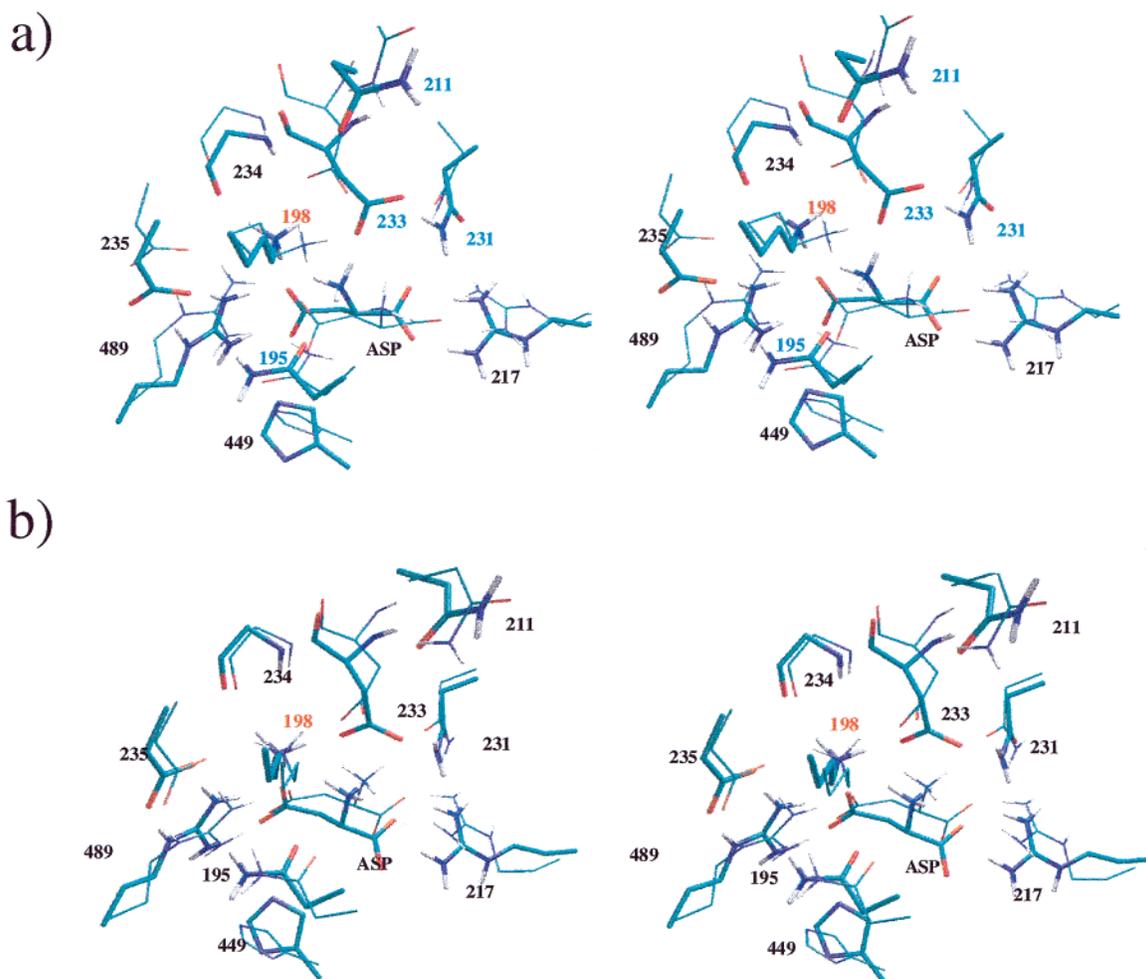
(52) Vlijmen, H. V.; Schaefer, M.; Karplus, M. *Proteins* **1998**, *33*, 145–158.

(53) Massova, I.; Kollman, P. *J. Am. Chem. Soc.* **1999**, *121*, 8133–8143.

(54) Mackerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W.; Roux, B.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorcikiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(55) Madura, J.; Briggs, J.; Wade, R.; Davis, M.; Luty, B.; Ilin, A.; Antosiewicz, J.; Gilson, M.; Baheri, B.; Scott, L.; Mccammon, J. *Comput. Phys. Commun.* **1995**, *91*, 57–95.

(56) Brooks, B.; Brucoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.



**Figure 2.** Closeups of the AspRS active site; divergent stereoviews. (a) Comparison between the mutant K198n:Asp structure obtained at the end of the K198  $\rightarrow$  K198n MDFE run 2 (thick lines), and the initial, native structure (thin lines). Residues whose interactions differ significantly between the two structures have colored labels. (b) Native complex AspRS:Asp. Thin: starting structure for MDFE run 1, which is close to the experimental structure. Thick: high energy final structure K198 $^\ddagger$  reached at the end of the “backward” run 3 (K198n  $\rightarrow$  K198), illustrating the structural hysteresis between the forward and backward MDFE runs. Figure produced with the programs VMD<sup>70</sup> and Raster3d.<sup>71</sup>

Lys198, and Gln231 interacts with the ligand main chain carboxylate (Figure 2a). As the charge on Lys198 is reduced, Asp233 rotates away from Lys198 and forms an interaction with Gln231. In backward runs 3 and 4, the Asp233:Gln231 interaction is present in the beginning (K198n state) and is maintained throughout the simulation, and it subsists in the K198 $^\ddagger$  state (Figure 2b). The final interaction, in K198 $^\ddagger$ , of Lys198 with Asp233 and with the ligand Asp side chain carboxylate is also weaker, compared to the native K198 structure. These structural differences define the high energy state K198 $^\ddagger$ . The resulting free energy hysteresis is partly cancelled by compensating contributions from residues Gln195, Gln211, Val234, and Arg489.

Presumably, the statistical weight of the high free energy structure K198 $^\ddagger$  reached at the end of runs 3 and 4 is very small, so that only the results from forward runs 1 and 2 are taken into account below.

The free energy cost to insert a negative charge on the Lys198 side chain is seen to be very large, similar to the value obtained earlier (95.1 kcal/mol) for inserting a positive charge on the Asp ligand.<sup>31,38</sup>

**3.2 Charge Insertion in the Asp Complex: Continuum Model.** We now show that a continuum model can serve to interpret the above results, provided static and relaxation steps are treated separately, with different protein dielectric constants.

The relaxation dielectric constant provides a measure of the polarizability of the active site.

The essential step is to determine the optimal protein dielectric constants; to do this, it is necessary to consider both the simple and compound charge insertion protocols described in the Methods section; see also ref 28. Indeed, the most useful criterion for choosing the dielectric constants is to require approximately the same overall free energy change  $\Delta G_{\text{Asp}}(\text{K198} \rightarrow \text{K198n})$  for all protocols, in agreement with the fundamental consistency conditions, eq 5. A second criterion is to obtain at least fair agreement with the overall MDFE free energy.

We first consider the static contributions, calculated at selected states. Relaxation contributions and the consistency between the static and relaxation terms are considered next.

**Static Free Energy.** To compute the static free energy, we require the equilibrium electrostatic potentials  $V_i$  at the charge insertion sites (the terminal atoms CE and NZ of the K198 side chain and the hydrogens bonded to them) in the reactant state (i.e., prior to the insertion). The potential  $V_i$  is created by the ligand and protein permanent and induced charges, including any charges present at the insertion sites in the reactant state. Notice that infinite interaction energies between the inserted charges and prior charges at the same sites cancel exactly when the reference free energy for the isolated Lys in a vacuum is subtracted; see the Methods section. The equilibrium positions

**Table 2.** Contributions from Selected Residues to the MDFE Free Energies<sup>a</sup>

Asp:AspRS								
residue	run					average	average	difference <sup>b</sup>
	1	2	3	4	5	1 + 2	3 + 4 + 5	
Asp233	101.9	97.1	86.5	84.9	87.6	99.5 ± 2.4	86.3 ± 1.1	13.2
ligand	54.7	52.2	32.5	44.0	29.0	53.5 ± 1.3	35.2 ± 6.4	18.3
Gln231	-0.8	1.1	-4.6	-4.4	-4.5	0.15 ± 0.95	-4.5 ± 0.1	4.7
Gln195	3.5	5.2	7.9	8.5	7.7	4.4 ± 0.9	8.0 ± 0.3	-3.6
Gln211	-1.9	2.0	4.7	4.1	5.3	0.05 ± 1.9	4.7 ± 0.5	-4.6
Val234	3.6	5.5	10.0	10.7	10.2	4.5 ± 0.9	10.3 ± 0.3	-5.8
Arg489	-80.1	-78.1	-70.0	-75.6	-66.6	-79.1 ± 1.0	-70.7 ± 3.7	-8.4
Asn:AspRS								
residue	run							difference
	1	2	3	4	5			2 - 4
Asp233	88.1	110.6	91.0	93.2	89.0			17.4
Glu235	93.2	91.5	82.6	77.9	94.8			13.6
Gln231	3.3	3.3	-1.7	-2.4	-3.6			5.7
His449	-6.0	0.0	-5.8	-4.1	0.2			4.2
Val234	2.7	2.6	6.0	5.7	5.9			-3.1
Gln211	-1.7	-1.5	3.0	2.8	2.8			-4.3
Gln195	13.5	8.2	12.1	13.4	13.3			-5.2
Tyr209	-2.3	-3.2	2.9	2.9	3.0			-6.1
Arg489	-77.0	-2.3	-65.8	-65.0	-66.1			-7.3

<sup>a</sup> All quantities in kcal/mol. The numbering of runs is the same as in Table 1. <sup>b</sup> Difference between the average values from the previous two columns.

**Table 3.** Continuum Electrostatic Free Energies for Charge Insertion into AspRS: Static Free Energies for Selected States<sup>a</sup>

state	ligand <sup>b</sup>	protein dielectric constant $\epsilon_p^s$	
		1	2
K198	Asp	171.6(12.1)	87.8(5.2)
K198n	Asp	39.5(7.3)	20.1(3.5)
midpoint <sup>c</sup>	Asp	111.5(11.6)	57.5(5.8)
K198	Asn	141.2(8.3)	71.2(4.2)
K198n	Asn	29.3(8.6)	
midpoint <sup>c</sup>	Asn	92.9(14.0)	

<sup>a</sup> In kcal/mol. For each state, calculations are done for 100–200 structures taken from an MD simulation of that state. Average values are reported (standard deviations in parentheses). <sup>b</sup> Results are reported for charge insertion into both the AspRS:Asp and AspRS:Asn complexes. <sup>c</sup> Averaged over structures taken from the midpoint K198/K198n ( $-e/2$  inserted charge) of forward and backward MDFE simulations.

of the protein and ligand atoms are generated by MD simulations of the corresponding state. The protein dielectric constant for this “static” step is denoted  $\epsilon_p^s$ .

In Table 3, we report the static free energy component for inserting a unit negative charge into selected “reactant” states, specifically the two endpoints (K198 and K198n) and the midpoint (K198/K198n). Results are reported for  $\epsilon_p^s = 1$  and 2 and are seen to scale approximately as  $1/\epsilon_p^s$ . With  $\epsilon_p^s = 1$ , the static free energy varies between 171.6 and 39.5 kcal/mol, depending on the state. The highest value (171.6 kcal/mol) is obtained at the K198 endpoint state. It arises from a large negative electrostatic potential at the location of the inserted charges, because of favorable interactions of nearby protein and ligand ionized groups with the positive Lys198 side chain. At the K198n endpoint, the static term is much smaller (39.5 kcal/mol), because the interaction between the neighboring charged residues and the neutralized Lys198 side chain is much weaker.

By considering the charging pathway of eq 3, we can already make a direct comparison between these static free energies and the total charging free energy obtained by MDFE simulations. Indeed, with this pathway, the contribution from the relaxation free energies (second term in eq 3) can be neglected (as shown

below). With  $\epsilon_p^s = 1$ , we then obtain a total charging free energy of 105.6 kcal/mol, reasonably close to the MDFE result of 94.4 kcal/mol (Table 1). Almost exact agreement with MDFE simulations would be obtained with  $\epsilon_p^s = 1.1$ . In contrast, choosing  $\epsilon_p^s \geq 2$  yields values of the static free energy that are much too small: for example, the value obtained in the K198 state (87.8 kcal/mol) is already smaller than the total free energy from MDFE simulations and would be further reduced upon adding the relaxation component. Thus, to obtain correct free energy values, a low dielectric constant,  $\epsilon_p^s \sim 1$ , is needed. This is in agreement with the results for a different charge insertion process in the AspRS active site.<sup>28</sup> As discussed in detail in ref 28, the low value of  $\epsilon_p^s$  is consistent with the use of molecular mechanics charges,<sup>54</sup> which are parametrized to reproduce the equilibrium electrostatic potential in the protein, without the need to include additional implicit polarization of the protein through  $\epsilon_p^s > 1$ .

Before proceeding to the relaxation free energies, we consider in more detail the structural origins of the static free energy. Selected group contributions are reported in Table 4. In the K198 state, Lys198 forms hydrogen bonds with the ligand Asp and the active site residues Asp233 and Glu235.<sup>38</sup> All these residues strongly oppose the neutralization of Lys198: the Asp ligand contributes  $\sim 31$  kcal/mol to the static term, Asp233 contributes 80.5, and Glu235 contributes 73.7 kcal/mol. Arg489 repels the positive Lys198 side chain and contributes  $-52.6$  kcal/mol to the static term. Arg217 makes a salt bridge with the ligand main chain carboxylate; it produces a positive potential that disfavors the K198 state. Other residues contribute smaller amounts.

At the K198n endpoint, the interaction between the neighboring charged residues and the neutralized Lys198 side chain is much weaker. A representative structure of the K198n:Asp state is shown in Figure 2a. Asp233 has rotated away from the neutral Lys198 and replaced its lost interaction by a stable hydrogen bond with Gln231. It also makes a strong interaction with the ammonium group of the ligand, which maintains the latter in the vicinity of the Lys198 side chain. At the same time, the ligand side chain carboxylate has shifted away from Lys198.

**Table 4.** Contributions from Individual Groups to the Static Free Energies

residue	K198:Asp	K198n:Asp	K198:Asn	K198n:Asn	av Asp <sup>b</sup>	av Asn <sup>c</sup>	difference <sup>d</sup>
ligand	31.0(6.9)	0.2(4.9)	-16.0(3.0)	-27.2(4.8)	15.6	-21.6	37.2
Asp233	80.5(7.9)	38.7(3.9)	92.4(4.7)	64.1(5.2)	59.6	78.3	-18.7
Glu235	73.7(6.2)	55.4(4.7)	70.9(14.5)	41.8(6.7)	64.6	56.4	8.2
Arg489	-52.6(2.8)	-50.4(3.6)	-45.2(5.9)	-38.1(6.6)	-51.5	-41.7	-9.8
Gln195	9.5(3.5)	5.9(2.9)	11.7(3.3)	6.1(4.1)	7.7	8.9	-1.2
Gln231	0.5(4.1)	-5.1(0.7)	4.6(0.8)	1.3(0.7)	-2.3	3.0	-5.3
Arg217	-5.8(1.0)	-5.8(0.8)	-3.9(0.6)	-6.6(1.0)	-5.8	-5.3	0.5
Lys198 <sup>e</sup>	25.9(0.4)	0.1(0.0)	25.0(1.0)	0.1(0.0)	13.0	12.5	0.5
Lys198 <sup>f</sup>	5.1(0.6)	4.6(0.7)	4.7(0.7)	4.6(0.6)	4.9	4.7	0.2
His449	-1.8(1.0)	-1.8(0.3)	0.7(0.6)	-4.6(1.7)	-1.8	-2.0	-0.2
total <sup>g</sup>	171.6	39.5	141.2	29.3			

<sup>a</sup> All quantities in kcal/mol. Standard deviations in parentheses. <sup>b</sup> Average over the endpoints in the Asp complex (evaluated from columns 2 and 3). <sup>c</sup> Average over the endpoints in the Asn complex (evaluated from columns 4 and 5). <sup>d</sup> Difference of average static terms (column 6 - column 7). <sup>e</sup> Contribution due to source charges at the insertion sites. <sup>f</sup> Contribution excluding source charges at the insertion sites. <sup>g</sup> Total static term values (from Table 3).

**Table 5.** Relaxation Free Energies for Selected States<sup>a</sup>

state	ligand	step <sup>b</sup>	protein dielectric constant $\epsilon_p^r$				
			2	3	4	6	8
		1	-38.2(0.2)	-51.3(0.4)	-58.3(0.4)	-65.2(0.5)	-68.8(0.6)
K198	Asp	2	-5.8(0.2)	-3.8(0.1)	-2.7(0.1)	-1.7(0.1)	-1.2(0.0)
K198n	Asp	2	-5.9(0.2)	-3.9(0.1)	-2.9(0.1)	-1.8(0.1)	-1.3(0.0)
K198	Asp	1 + 2	-44.0(0.2)	-55.1(0.1)	-61.0(0.1)	-66.9(0.1)	-70.0(0.0)
K198n	Asp	1 + 2	-44.1(0.2)	-55.2(0.1)	-61.2(0.1)	-67.0(0.1)	-70.1(0.0)
K198	Asn	2	-6.1(0.2)	-3.9(0.1)	-2.9(0.1)	-1.8(0.1)	-1.3(0.0)
K198n	Asn	2	-6.2(0.3)	-4.0(0.2)	-2.9(0.1)	-1.8(0.1)	-1.3(0.1)
K198	Asn	1 + 2	-44.3(0.2)	-55.2(0.1)	-61.2(0.1)	-67.0(0.1)	-70.1(0.0)
K198n	Asn	1 + 2	-44.4(0.3)	-55.3(0.2)	-61.2(0.1)	-67.0(0.1)	-70.1(0.1)

<sup>a</sup> In kcal/mol. Average values (standard deviations), as in Table 3. <sup>b</sup> In step 1, the perturbing charges are introduced into an infinite protein medium; in step 2, the outer region is changed into solvent (see Methods section).

These conformational rearrangements lead to cancellations between the potentials of the ligand ammonium and carboxylate groups and a small ( $\sim 0$ ) total static term due to the Asp ligand. Asp233 and Glu235 still provide disfavoring (albeit smaller) contributions of 38.7 and 55.4 kcal/mol, respectively (see Table 4). Gln231 rotates and hydrogen-bonds to the negative Asp233; in this orientation, it disfavors a positive Lys198 by  $-5.1$  kcal/mol.

**Relaxation Free Energy.** Relaxation free energies are calculated for insertion of a unit charge into the same states considered above. The relaxation free energy is calculated in two stages, as explained in the Methods section and in ref 28. The first stage corresponds to the transfer of the insertion charges from vacuum to a cavity with the shape of the Lys198 side chain and a dielectric constant of 1, which is embedded in an infinite “protein” medium having a dielectric  $\epsilon_p^r$ . For  $\epsilon_p^r \geq 2$ , this term gives the dominant contribution to the relaxation (see Table 5). It increases with  $\epsilon_p^r$  approximately as  $1 - 1/\epsilon_p^r$ , a dependency also found earlier for the Asp  $\rightarrow$  Asn transformation.<sup>28</sup> The second stage, namely, the transformation of the outer (solvent) dielectric from  $\epsilon_p^r$  to  $\epsilon_w = 80$ , yields a smaller free energy contribution. This is because the Lys198 side chain is not accessible to the solvent in any of the states considered (K198, K198n, or the midpoint K198/K198n) and interacts only weakly with polarization charge on the protein-solvent dielectric boundary. The total relaxation free energy ranges from  $-44.0$  kcal/mol with  $\epsilon_p^r = 2$  to  $-70.0$  kcal/mol with  $\epsilon_p^r = 8$ . The observed similarity between the relaxation free energies in the reactant and product states is in accord with linear response and is manifested here for all values of  $\epsilon_p^r$ .

**Total Charging Free Energy and Comparison between Pathways.** In Table 6, we summarize the values of the static and relaxation free energies and their mutual consistency for

**Table 6.** Relations between Static and Relaxation Free Energies<sup>a,b</sup>

$\epsilon_p^r$	$1/2(\Delta G_r^{K198} + \Delta G_r^{K198n})$	$1/2(\Delta G_s^{K198n} - \Delta G_s^{K198})$	$\epsilon_p^s$
<b>AspRS:Asp Complex</b>			
3	-55.2	<b>-66.1</b>	1
4	-61.0	-33.9	2
6	<b>-67.0</b>		
8	-70.0		
<b>AspRS:Asn Complex</b>			
2	-44.4	<b>-56.0</b>	1
3	<b>-55.3</b>		
4	-61.0		
6	-67.0		

<sup>a</sup> In kcal/mol. <sup>b</sup> According to the linear response relations (eq 5),  $\Delta G_r^{\text{react}} = \Delta G_r^{\text{prod}}$  and  $\Delta G_s^{\text{prod}} - \Delta G_s^{\text{react}} = \Delta G_r^{\text{prod}} + \Delta G_r^{\text{react}}$ . Values satisfying these relationships approximately are in bold face. Here, K198 and K198n play the role of reactant and product states, respectively.

various choices of  $\epsilon_p^s$  and  $\epsilon_p^r$ . In Table 7, we add the static and relaxation contributions to obtain the total free energy  $\Delta G_{\text{Asp}}$  (K198  $\rightarrow$  K198n) for different combinations of dielectric constants and various simple and compound pathways.

We showed above that a value of  $\epsilon_p^s = 1$  leads to a charging free energy in reasonable agreement with MDFE simulations for the compound charging pathway of eq 3. With this value of  $\epsilon_p^s$ , the best consistency between static and relaxation free energies (in the sense of eq 5) corresponds to  $\epsilon_p^r = 6$ . This leads to a total charging free energy of 104.7 kcal/mol for the simple K198  $\rightarrow$  K198n pathway, in fair agreement with the MDFE simulation result, 94.4 kcal/mol (Table 1). The combinations  $(\epsilon_p^s, \epsilon_p^r) = (1, 4-8)$  give a range of values, 100.7–110.6 kcal/mol (Table 7).

Compound pathways with  $(\epsilon_p^s, \epsilon_p^r) = (1, 6)$  lead to similar results. In the pathway K198  $\leftarrow$  K198/K198n  $\rightarrow$  K198n,  $\pm 1/2e$  charges are inserted into the midpoint state K198/K198n, and

**Table 7.** Total K198  $\rightarrow$  K198n Free Energy from Continuum Model. Estimation from Different Dielectric Constants and Pathways<sup>a</sup>

AspRS:Asp Complex					
path <sup>b</sup>	static	$\epsilon_p^s$	relaxation	$\epsilon_p^r$	sum
K198 $\rightarrow$ K198n	171.6	1	-44.0	2	127.6
			-61.0	4	110.6
			-66.9	6	<b>104.7</b>
			-70.0	8	101.6
			87.8	2	-44.0
K198 $\leftarrow$ K198n	39.5	1	61.2	4	100.7
			67.0	6	<b>106.5</b>
			68.1	8	107.6
K198 $\leftarrow$ K198/K198n $\rightarrow$ K198n	111.5	1			<b>111.5</b>
K198 $\rightarrow$ K198/K198n $\leftarrow$ K198n	105.6	1	$\sim 0$	6	<b>105.6</b>
quartersteps	108.6	1	$\sim 0$	6	<b>108.6</b>
average <sup>c</sup>					<b>107.4 <math>\pm</math> 2.4</b>
MD free energy <sup>d</sup>					94.4
AspRS:Asn Complex					
path <sup>b</sup>	static	$\epsilon_p^s$	relaxation	$\epsilon_p^r$	sum
K198 $\rightarrow$ K198n	141.2	1	-55.2	3	<b>86.0</b>
			-61.1	4	80.1
			-55.2	3	20.0
			29.3	1	44.4
K198 $\leftarrow$ K198n	29.3	1	55.3	3	<b>84.6</b>
			61.1	4	90.4
			67.0	6	96.3
			92.9	1	
K198 $\rightarrow$ K198/K198n $\leftarrow$ K198n	85.3	1	$\sim 0$	3	<b>85.3</b>
quartersteps	88.1	1	$\sim 0$	3	<b>88.6</b>
average <sup>c</sup>					<b>87.1 <math>\pm</math> 2.3</b>
MD free energy <sup>d</sup>					77.2

<sup>a</sup> In kcal/mol. <sup>b</sup> The initial state corresponds to the tail of the arrows. K198/K198n is the midpoint state. <sup>c</sup> Bold face numbers are used to obtain the average values. <sup>d</sup> Free energy calculated by molecular dynamics free energy (MDFE) simulations (see Table 1).

the perturbation is performed toward the left and right endpoints. As discussed after eq 4 and demonstrated in ref 28, the relaxation terms cancel out exactly in this pathway. In the pathway K198  $\rightarrow$  K198/K198n  $\leftarrow$  K198n, the charge is inserted in  $\mp^{1/2}e$  steps at the left and right endpoints. In this case, the total free energy includes a general contribution that is proportional to the difference of reactant and product relaxation terms (see eq 3); these terms are approximately equal here and cancel out in the difference. The resulting average (105.6) agrees well with the value along the other two-step pathway (111.5). Notice that this particular pathway coincides with the one used by Muegge et al.<sup>12</sup> for charge insertion in cytochrome *c*. A more complicated path utilizes quartersteps, in the scheme K198  $\rightarrow$  A  $\leftarrow$  K198/K198n  $\rightarrow$  B  $\leftarrow$  K198n. At each step, a  $\pm^{1/4}e$  charge is inserted. Again, the value (108.6) is in good agreement with the midpoint value.

Averaging over the various simple and compound pathways, the combination  $(\epsilon_p^s, \epsilon_p^r) = (1, 6)$  gives a charging free energy of  $107.4 \pm 2.4$  kcal/mol, in fair agreement with the MDFE result, 94.4 kcal/mol. When the difference between the charging free energy in the AspRS:Asp and AspRS:Asn complexes is taken (below), the agreement is significantly better. Notice that if we were to impose  $\epsilon_p^s = \epsilon_p^r$ , as in most implementations of continuum models, the consistency relations in eq 5 would require a dielectric of about  $\epsilon_p^s = \epsilon_p^r = 1.6$ , giving a total charging free energy of about 65 kcal/mol, in very poor agreement with the MDFE value.

In summary, to obtain a consistent and accurate continuum model, different values of  $\epsilon_p^s$  and  $\epsilon_p^r$  are needed. The charge

insertion studied here then reports a protein dielectric constant of about 6 for relaxation in the AspRS active site.

**3.3 Charge Insertion in the AspRS:Asn Complex: MDFE Simulations.** We now discuss the results for the charge insertion K198  $\rightarrow$  K198n in the AspRS:Asn complex, beginning with the molecular dynamics free energy simulations. Five runs were performed; results are listed in Table 1, and selected residue free energy components are included in Table 2. The starting points for forward runs 1 and 2 and backward runs 3 and 4 were different snapshots from long (several hundred ps) simulations of the K198 and K198n end states. Forward run 5 was started after a 200 ps simulation at the end of backward run 3.

The free energies obtained have a spread of 13.7 kcal/mol, which is analogous to that of the AspRS:Asp complex. Forward run 1 has the largest free energy (96.8); backward runs 3 and 4 have the smallest values (83.1 and 84.4), and the other forward values are intermediate.

An examination of the structures and an analysis of the free energy components associated with individual residues show that the differences among runs arise from many structural groups. To show this, we include in Table 2 the residues that contribute the most to the difference between the highest and lowest free energy runs 2 and 4. The residue components for the various runs are in columns 2–6, and the difference between runs 2 and 4 is in column 7. Residues Asp233, Glu235, Gln231, and His449 contribute the most and are partially cancelled by residues Arg489, Tyr209, Gln195, Gln211, and Val234. Among these residues, the components of Asp233, His449, Gln195, and Arg489 vary also considerably between forward runs 1 and 2; their total contribution to the difference (run 1 minus run 2) is 17.6 kcal/mol. Numerous other residues (not shown) reduce this contribution to the observed total difference of 6.3 kcal/mol between runs 1 and 2 (see Table 1). Runs 1 and 4 differ also in a number of components (e.g., Glu235, Gln231, Arg489).

The two backward runs 3 and 4 are very similar in total values and individual residue components. Forward run 5 starts at the end of run 3 and proceeds in the opposite direction. Compared to runs 3 and 4, it has a contribution due to Glu235 that is larger by 12–15 kcal/mol. Cancellations due to other residues (not shown) lead to a smaller (3–4 kcal/mol) total difference between this and the backward runs.

Unlike the AspRS:Asp complex, the differences among the various runs originate from a large number of residues, indicating that a large variety of structures is sampled. For this reason, we average the free energy results over all runs and obtain a K198  $\rightarrow$  K198n free energy of  $77.2 \pm 5.5$  kcal/mol, compared to 94.4 kcal/mol with the Asp ligand (see above). The smaller free energy cost is consistent with the absence of a negative charge on the ligand (Asn) in the present complex. If we were to average over only forward runs 1–2, we would obtain 82.5 kcal/mol.

**3.4 Charge Insertion in the AspRS:Asn Complex: Continuum Model. Static and Relaxation Contributions to the Charging Free Energy.** Here, as in the AspRS:Asp complex, the static component  $\Delta G_s$  of the free energy depends strongly on the reactant state and the dielectric constant  $\epsilon_p^s$  (Tables 3, 6). We limit our discussion to the case  $\epsilon_p^s = 1$ , because this value was shown above to be the most physically appropriate. The static free energy in the K198 state (141.2 kcal/mol) is smaller than the AspRS:Asp result. In the MD simulations of this state, the Asn ligand's side chain NH<sub>2</sub> group is repelled by Arg489, and the ligand is displaced and rotated compared to the cognate ligand Asp. The ligand–Lys198 interaction is lost,

and a new stable interaction is formed between Asp233 and the ligand ammonium, keeping the latter in the vicinity of Lys198. The proximity of the ligand ammonium to Lys198 is unfavorable and largely accounts for the preference of the Asn ligand for a neutral Lys198, demonstrated by its negative contribution to the static term ( $-16.0$  kcal/mol). Asp233 and Glu235 make strong contributions ( $92.4$  and  $70.9$  kcal/mol) to  $\Delta G_s$  that oppose a neutral Lys198. These residues dominate the overall static free energy.

The endpoint K198n corresponds to the smallest static term ( $29.3$  kcal/mol), still favoring a charged Lys198. The ligand provides an even more negative contribution of  $-27.2$  kcal/mol in favor of a neutral Lys198. Asp233 and Glu235 contribute  $64.1$  and  $41.8$  kcal/mol, favoring a charged Lys198. The effect of Glu235 is mostly cancelled by its salt bridge partner Arg489, yielding a total of  $3.7$  kcal/mol for this pair, in close agreement with its contribution in the K198n:Asp complex ( $5.0$  kcal/mol). Other residues make smaller contributions.

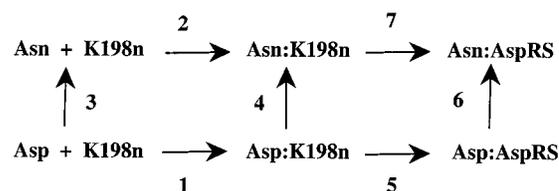
At the two endpoints, K198 and K198n, the static terms overestimate or underestimate the total free energy, an effect that is corrected upon adding the negative relaxation component. The relaxation free energy is shown in Table 5 for selected reactant states and various choices of  $\epsilon_p^r$ . The relaxation free energy is approximately the same in the two endpoint states (K198 and K198n), as expected from linear response, and is very similar to the values for the AspRS:Asp complex.

The consistency between the static and relaxation components is shown in Table 6 for various choices of  $\epsilon_p^s$  and  $\epsilon_p^r$ . The best consistency (bold face) corresponds to a dielectric of 1 for the static term, as in the AspRS:Asp complex, and a somewhat smaller relaxation dielectric of 3 (compared to 6 for the Asp complex).

**Total Charging Free Energy and Comparison between Pathways.** The total free energy for charge insertion in the AspRS:Asn complex is given in Table 7 for different combinations of dielectric constants and various simple and compound pathways. The combination  $(\epsilon_p^s, \epsilon_p^r) = (1, 4-8)$  yields free energies in the range  $70-99$  kcal/mol, that is, with a somewhat larger spread than in the Asp complex. The optimum pair (1, 3) gives  $86.0$  or  $84.6$  kcal/mol, and an average over simple and compound pathways yields a value of  $87.1$  kcal/mol. This is in fair agreement with the average value from MDFE,  $77.2$  kcal/mol.

The difference in the average charging free energies for the AspRS:Asp and AspRS:Asn complexes is  $20.3$  kcal/mol. It originates mainly from the static terms, because the relaxation terms are similar in the two complexes and cancel out. Nevertheless, the dielectric relaxation of the protein is incorporated explicitly in this difference through the use of different protein structures for each state. The double free energy difference is in good agreement with MDFE, which yields  $94.4-77.2 = 17.2$  kcal/mol.

Notice that the continuum values could easily be adjusted to produce results in still closer agreement with MDFE, by merely setting the static dielectric constant  $\epsilon_p^s$  to a value slightly greater than 1 and adjusting  $\epsilon_p^r$  accordingly (to maintain consistency in the sense of eq 5). For example, with the observed, approximate  $\sim 1/\epsilon_p^s$  dependence of the static terms, the value  $\epsilon_p^s = 1.12$  leads to overall continuum free energies of  $95.6$  and  $77.8$  kcal/mol for AspRS:Asp and AspRS:Asn, compared to  $94.4$  and  $77.2$  kcal/mol with MDFE. The resulting difference ( $17.8$  kcal/mol) is also in better agreement with the MDFE prediction ( $17.2$  kcal/mol). However, an exact fit of the molecular dynamics results is not the object of the present



**Figure 3.** Thermodynamic cycle to analyze ligand binding to modified AspRS (K198n). The binding free energy difference between Asp and Asn is  $\Delta G_1 - \Delta G_2$  (indices refer to leg numbers in the figure). It is also equal to  $\Delta G_3 - \Delta G_4 = \Delta G_3 - (\Delta G_5 + \Delta G_6 - \Delta G_7)$ .  $\Delta G_5$  and  $\Delta G_7$  were calculated in this work using both MDFE and a continuum model;  $\Delta G_3 - \Delta G_6$  was calculated in ref 38 using MDFE and in ref 39 using a continuum model.

continuum model. Rather, it is to obtain a simple, phenomenological description of the dielectric relaxation in response to charge insertion. An improved fit would not change the main result of this work: a protein dielectric constant  $\epsilon_p^r$  of 3–6 is needed to reproduce the dielectric relaxation found by molecular dynamics.

The  $20.3$  kcal/mol (continuum model) or  $17.2$  kcal/mol (MDFE) difference between the charging free energies for Asp:AspRS and Asn:AspRS corresponds to a substantially stronger Asn binding to the K198n protein, compared to the native K198 protein (see the thermodynamic cycle in Figure 3). Combining the MDFE data from this work and from ref 38, the binding free energy difference is reduced from  $15.3$  kcal/mol for native AspRS (favoring Asp binding) to  $-1.9$  kcal/mol for the mutant K198n protein. Thus, deprotonation of Lys198 (or its substitution by a neutral residue of about the same size, such as leucine) should strongly enhance the binding of Asn, making it an efficient inhibitor of the mutant AspRS.

#### 4. Concluding Discussion

The complexity of dielectric relaxation in proteins makes its experimental characterization difficult, and so theoretical approaches are valuable. The comparison of detailed atomistic simulations with simpler models such as a dielectric continuum model is especially useful for obtaining qualitative insights. A continuum description of a protein is neither rigorous nor unique, as discussed by many authors.<sup>1,10,12,27,41,47,57-59</sup> However, its simplicity and physical transparency are precisely what make it useful for a qualitative interpretation of the microscopic protein dynamics.

Here, we have analyzed a charge insertion process that models deprotonation or mutation of an important side chain in the active site of the enzyme aspartyl-tRNA synthetase (AspRS). Two AspRS:ligand complexes were treated; charge insertion on the ligand itself was studied earlier.<sup>28</sup> By considering two ligands and two charge insertion sites, we get a sense of the spatial variation of the dielectric properties and of the robustness of the continuum interpretation. Dielectric relaxation was found to involve both ligand and side chain rearrangements in the active site and to account for a large part of the overall charging free energy. A moderate protein dielectric constant of 3–6 (depending on the insertion site and the nature of the ligand) is needed to describe relaxation in response to the charge insertion. This indicates that the local protein polarizability in the AspRS active site is not much greater than that expected at nonspecific positions in a protein interior.<sup>40,41</sup> Note that a larger local

(57) Warshel, A.; Aqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 268–298.

(58) Simonson, T.; Perahia, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082–1086.

(59) Krishtalik, L.; Kuznetsov, A.; Mertz, E. *Biophys. J.* **1996**, *70*, A225.

polarizability could occur in other sites and other systems.<sup>26</sup> The time dependence of dielectric relaxation<sup>11,17,60</sup> was not analyzed here, although it is expected to play a role in the kinetics of actual charge-transfer reactions occurring in AspRS.

Two theoretical advances made the present analysis possible. The first allowed long-range interactions to be included in the molecular dynamics free energy simulations. Accurate treatments of long-range electrostatics for biomolecular simulations have been available for several years, particularly lattice methods such as Ewald summation<sup>61,62</sup> and multipole methods.<sup>63–65</sup> Lattice methods are expensive, because the protein must be fully immersed in explicit water, and they have seldom been used so far in free energy simulations of proteins. Langevin dipoles provide another approach.<sup>7</sup> An alternative, continuum reaction field method for protein free energy simulations was first proposed in ref 31 and found here to be accurate and efficient; see also ref 32. Of course, if the charging free energy is available from experiment, the MDFE calculations are not strictly necessary (although MD simulations are still useful to provide structural models for the FDP continuum calculations).

The second advance was the recognition that, with a continuum model, the two-step charging procedure of Marcus could be used to “decouple” the description of the equilibrium potential (static step), which depends strongly on the choice of force field parameters (including the charge set), and the dielectric relaxation free energy, which is essentially independent of the charge set. By calculating the charging free energy along different simple and compound pathways, we can then focus on either the static or the relaxation component of the free energy, or both. Thus, some pathways are sensitive to  $\epsilon_p^s$  (e.g., those that use the reactant and product states with equal “weights”; eq 3); others are sensitive to both  $\epsilon_p^s$  and  $\epsilon_p^r$  (e.g., those using only the reactant structure; eq 1 with  $\lambda = 0$ ).

The molecular mechanics charge set used here was parametrized for MD simulations with a dielectric constant of 1;<sup>54</sup> that is, it should reproduce the equilibrium electrostatic potential without the need to include additional polarization charge in the protein through  $\epsilon_p^s > 1$ . On the other hand, dielectric relaxation of the protein must be described by a dielectric  $\epsilon_p^r > 1$ . The values of  $\epsilon_p^s \approx 1$  and  $\epsilon_p^r = 3–6$  obtained for AspRS are consistent with this picture. The  $\epsilon_p^r$  range is similar, as it should be, to the dielectric constants calculated from the dipole fluctuations of several proteins in MD simulations.<sup>40,66</sup> It is expected that the same situation will arise for other systems and other charge sets: different protein dielectric constants will be needed to describe static fields ( $\epsilon_p^s$ ) and field shifts induced

by a perturbing charge ( $\epsilon_p^r$ ). This indicates that in some current applications of continuum models, the charge set and dielectric constants may not be mutually consistent. This could affect previous studies of the polarity and polarizability of protein interiors with continuum models; for example, refs 21, 24, 52, and 67.

In principle, the static and relaxation components of the charging free energy are both measurable under certain conditions. Upon photoexcitation of a protein-bound chromophore, the redistribution of electron density can be modeled as a rearrangement of point charges. The corresponding static free energy is then closely related to the Stark shift of the chromophore absorption band, relative to its absorption band in a nonpolar medium, while the relaxation free energy is related to the Stokes shift of the emission band. The high-frequency dielectric constant is usually thought to be appropriate for the former, and the zero-frequency dielectric constant for the latter. Thus, the use of different dielectric constants for static and relaxation steps is rooted in classic ideas from spectroscopy and electron transfer theory.<sup>59,46</sup>

The protein polarizability observed here in the active site of AspRS is moderate, corresponding to a relaxation dielectric constant of only 3–6. This is slightly higher than the polarizability deduced from Stokes shift measurements for a probe bound in the active site of the enzyme chymotrypsin in dehydrated films.<sup>19</sup> Dielectric dispersion by dry protein powders also gives low dielectric constants.<sup>68</sup> The shielding of charge–charge interactions between chromophores in the photosynthetic reaction center, estimated from their Stark shifts in different oxidation states, corresponds to an average protein dielectric constant of 2–4.<sup>16</sup> Molecular dynamics simulations give values of 1–4 for the dielectric constant within protein interiors<sup>40</sup> (though much larger values are obtained for flexible surface regions<sup>40,66</sup>). A somewhat larger value of 8 was obtained for the active site region of trypsin.<sup>26</sup>

The same total relaxation free energy for charge insertion in the AspRS active site, including both protein and solvent contributions, would also be obtained in a homogeneous medium with a dielectric constant of about 4–7. Thus, the contribution of solvent to the polarizability in this region is small, raising the average dielectric constant by only one unit. This “effective” average dielectric constant is similar to that of chloroform (4.8) or acetic acid (6.2), an order of magnitude smaller than that of bulk water. The moderate polarizability is in contrast to the very large polarity of the active site,<sup>69</sup> measured, for example, by the large static free energy component for charge insertion on either Lys198 or the Asp ligand. The interplay between a large polarity and a moderate polarizability is expected to be important for efficient charge transfer and catalysis.

JA010716+

(60) Song, X.; Chandler, D.; Marcus, R. *J. Phys. Chem.* **1996**, *100*, 11954–11959.

(61) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(62) Darden, T. In *Computational Biochemistry & Biophysics*; Becker, O., Mackerell, A., Jr., Roux, B., Watanabe, M., Eds. Marcel Dekker: New York, 2001; Chapter 4.

(63) Board, J. A., Jr.; Causey, J. W.; Leathrum, J. F., Jr.; Windemuth, A.; Schulten, K. *Chem. Phys. Lett.* **1992**, *198*, 89.

(64) Stote, R.; States, D.; Karplus, M. *J. Chim. Phys.* **1991**, *88*, 2419–2433.

(65) Lee, F.; Warshel, A. *J. Chem. Phys.* **1992**, *95*, 4366–4377.

(66) Pitera, J.; Falta, M.; Van Gunsteren, W. *Biophys. J.* **2001**, *80*, 2546–2555.

(67) Kim, Y.; Kim, H.; Lee, Y. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 99–103.

(68) Bone, S.; Pethig, R. *J. Mol. Biol.* **1985**, *181*, 323–326.

(69) Warshel, A. *Nature* **1987**, *330*, 15–17.

(70) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

(71) Merritt, E.; Murphy, M. *Acta Crystallogr. D* **1994**, *50*, 869–873.