Continuum Treatment of Long-Range Interactions in Free Energy Calculations. Application to Protein−Ligand Binding.

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · OCTOBER 1997

Impact Factor: 3.3 · DOI: 10.1021/jp9711499

3 AUTHORS, INCLUDING:

Thomas Simonson
École Polytechnique

116 PUBLICATIONS **20,665** CITATIONS

SEE PROFILE

Continuum Treatment of Long-Range Interactions in Free Energy Calculations. Application to Protein−Ligand Binding.

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

Laboratoire de Biologie Structurale (CNRS), IGBMC, 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, and Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received: April 1, 1997; In Final Form: June 11, 1997[⊗]

A method is proposed to include long-range electrostatic interactions in free energy calculations that involve the creation or deletion of net charges in a macromolecule. The vicinity of the mutation site is treated microscopically, while distant bulk solvent is treated macroscopically. A three-step mutation pathway is used. First, the mutation is introduced with a molecular dynamics simulation for the macromolecule, solvated by a limited number of explicit water molecules and surrounded by vacuum. Selected charges may be reduced during this step to mimic the effect of bulk solvent on the dynamics and to obtain a simulation in which the structures sampled are correct. The full effect of bulk solvent is accounted for in the next two steps. In the second step, the reduced charges are increased to their original values and the corresponding free energy change is obtained from continuum electrostatics. In the third step, the system is transferred into bulk solvent, modeled as a dielectric continuum, and the transfer free energy is obtained from continuum electrostatics. A potential-based charge scaling can be used in step I for the selected charges, which reduces each one in proportion to the screening by bulk solvent of its potential at the mutation site. With this method, the *a priori* scaling of a charged group in step I is formally equivalent to its solvation by bulk solvent in step III. Thus, for a given charged group, *a priori* scaling or *a posteriori* bulk solvation should give similar results. The method is illustrated by a calculation of the free energy change associated with the mutation of an aspartate ligand into asparagine in the active site of aspartyl-tRNA synthetase, a process that removes a negative charge. The results of five free energy runs with three different charge scaling schemes are in good agreement. This indicates that the method is robust with respect to implementation details and that the continuum approximation in steps II and III is valid for this case.

1. Introduction

Electrostatic interactions are central to protein structure and function. They play an important role, for example, in enzyme transition state stabilization, in oxidation−reduction processes, in the pH-dependent ionization of acidic and basic groups, and in molecular recognition. Computer simulations have the potential to provide a quantitative microscopic description of these interactions. Several techniques have been used for this purpose.

The simplest method is the macroscopic continuum model, which usually treats the macromolecule as a single low-dielectric medium with embedded fixed charges, surrounded by a high-dielectric medium representing the solvent.[1,2] Successful applications to charge creation or separation in proteins have been reported.[3−5] The most detailed studies involve pK_a calculations of ionizable groups.[6−8] However, the simple continuum model does not adequately account for dielectric inhomogeneity or nonlinearities, *e.g.*, at protein−water interfaces;[9−11] efforts to incorporate these effect are underway in several groups.[8,12−15] The Protein Dipole Langevin Dipole model[16] provides an alternative, with essential similarities to the continuum model, and at least some of its limitations. It relies on an atomic point polarizability model of the protein, whose dielectric response has been accurately reproduced by a continuum model for at least two proteins.[9,17]

Microscopic models can provide a much more detailed description. Free energy simulations, in particular, have been used to study electrostatic interactions.[18−22] However, the technical difficulties are considerably greater; *e.g.*, to be successful, free energy simulations must treat long-range interactions accurately. Molecular dynamics simulations of a macromolecule fully solvated in a box of water, including long-range interactions through Ewald summation, have become feasible.[23] Simulations in spherical water droplets have also been done, with a continuum treatment of the reaction field produced by surrounding bulk water.[24−26] However these approaches include the entire protein along with several explicit solvation layers, and so they remain very expensive even for the smallest proteins. Stochastic boundary simulations[27,28] have been performed on limited portions of a protein, with a multipole model for the more distant protein charges;[29] effective scaling of protein charges has also been used to mimic screening by bulk solvent.[30,31] Warshel and co-workers have used models of limited portions of proteins with explicit solvent, surrounded by networks of Langevin dipoles, to study charge creation or charge separation in several enzymes;[22,32,33] the Langevin dipoles follow a simplified dynamics, and are themselves embedded in a surrounding dielectric continuum. Recently, methods have been developed to incorporate macroscopic continuum treatments of solvent regions into molecular dynamics simulations.[34−36] By these methods, a subset of a protein, or the entire protein,

* Corresponding authors.
† CNRS.
‡ Université Louis Pasteur.
§ Harvard University.
⊗ Abstract published in *Advance ACS Abstracts,* August 15, 1997.

S1089-5647(97)01149-8 CCC: $14.00

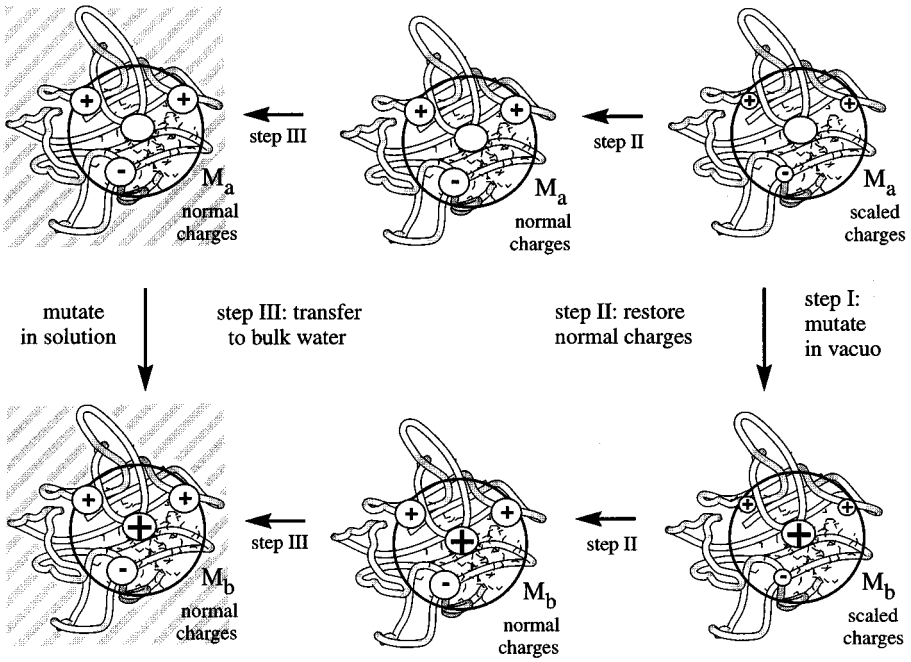

Figure 1. Three-step thermodynamic cycle to compute the $M_a \rightarrow M_b$ free energy difference in bulk solution. Step I: the free energy difference is computed for a microscopic subsystem *in vacuo*, with scaling of selected charges (schematized above as small spheres). In the present application, a spherical region of the protein and 384 explicit waters are treated by molecular dynamics; the remainder of the protein is fixed; the mutation adds a single charge near the center. In the cartoon above, the protein is shown as a coil, explicit waters are shown, and the spherical region is indicated by a circle. Step II: the scaled charges are restored to their normal values (schematized above as larger spheres), for both M_a and M_b, and the corresponding free energy changes are calculated with a Poisson−Boltzmann (PB) continuum model. Step III: M_a and M_b are both transferred into bulk solvent, and the corresponding free energy changes are calculated with a continuum model.

along with explicit water molecules, can be simulated in microscopic detail, with a surrounding high-dielectric continuum. However, most of these methods make the assumption that the dielectric medium outside the microscopic region is homogeneous. Thus if portions of the protein are not explicitly included in the microscopic region, they are effectively neglected and replaced by bulk solvent.

A simple hybrid approach has been suggested recently,[37,38] similar in spirit to the Born correction commonly used in free energy simulations of small solutes.[39] Free energies are first calculated for a spatially limited model, *e.g.*, a solute along with a limited number of explicit solvent molecules, using microscopic simulations such as molecular dynamics. The finite model is then transferred into bulk solvent, and the transfer free energy is calculated with a continuum model. This approach was tested on small solutes,[37] and a simplified version was applied to a pK_a calculation in a protein.[38] We present here a modification of this approach and a detailed application to a protein−ligand binding problem. In our implementation, charge screening by bulk solvent is included approximately in the microscopic simulations by scaling factors applied to selected charged protein groups.[30] Because bulk water is not included here, the method is computationally efficient. To correct for the absence of bulk solvent, the natural charges are then restored, and the finite model is transferred into a dielectric continuum bulk solvent. The dielectric inhomogeneity of the macromolecule−solvent system is taken into account through the numerical finite-difference solution of the Poisson−Boltzmann equation.

The method is illustrated by a calculation of the free energy difference between two protein−ligand complexes: aspartyl-tRNA synthetase complexed with its substrate aspartic acid and with the substrate analogue asparagine. This calculation represents one leg of a thermodynamic cycle, of a type commonly used to study protein−ligand binding.[40] The other leg is the alchemical mutation of aspartic acid into asparagine in solution. The difference between the free energies for these two legs is equal to the difference in binding free energies between these two ligands. Calculations of the mutation free energy in solution will be reported elsewhere, along with a detailed analysis of the present simulations of the protein−ligand complexes.[41]

The method is described in detail in the next section. The third section describes the computational details; the fourth section describes the results; and the fifth section is a discussion.

2. Theory

2.1. Three-Step Thermodynamic Cycle. We give here an overview of the method and the three-step thermodynamic cycle upon which it is based. In the following sections we consider each step in detail.

Consider two macromolecules in solution, M_a and M_b, which differ in their net charges. To calculate the free energy difference between them we use a three-step approach (Figure 1), as follows. (I) In step I we treat the macromolecules with a detailed microscopic model, with or without a large but finite number of explicit water molecules. This finite model, in which bulk solvent is not included, is surrounded by vacuum. We calculate the free energy difference between M_a and M_b using molecular dynamics and a free energy perturbation approach.[19,20] These finite models, in the absence of bulk solvent, could have significant structural and dynamic distortions, compared to the fully solvated systems. Ionized protein side chains near the surface of such a model, for example, could have very different orientations with and without bulk solvent. To reduce such distortions, the charges of selected ionized groups are reduced, to roughly mimic screening by bulk solvent. It is important to note that we do not seek to model bulk screening exactly at the present stage, because this is done in the subsequent steps. (II) In step II, we remove the charge scaling, restoring all charges to their normal values. The free energy for this step is calculated

with a Poisson−Boltzmann continuum model, in which the system (macromolecule + explicit solvent) is viewed simply as a set of fixed charges in a vacuum. The calculation is performed for both M_a and M_b. (III) In step III, we immerse the system in a high dielectric continuum, representing bulk solvent. The free energy for this transfer step is calculated from a continuum model. The calculation is performed for both M_a and M_b.

This last step is reminiscent of the Born correction for cutoff effects, widely used when calculating the difference in solvation free energies between a neutral and a charged solute in solution, for a simulation system with periodic boundary conditions and an electrostatic cutoff.[42] In our method, the continuum model is used not as a correction for cutoff effects, but to model the transfer of a finite system into bulk solution. This can be done exactly within the framework of continuum electrostatics.

The charge scaling (step I) can be done in several different ways. By comparing several protocols, the robustness and reliability of the method can be tested. The scaling factors do not have to model exactly the effect of solvent screening, as long as they reduce the structural and dynamic distortions produced in the model by the absence of bulk solvent. Indeed, if structural distortions in the vacuum system are too large, the continuum model in step III cannot accurately account for the transfer into bulk solvent. For proteins, one protocol is to apply scaling factors to exposed, unsolvated, ionized side chains.[30,31] In practical applications, M_a and M_b will often differ by a single side chain mutation, as in the present application. Reasonable scaling factors α_i, for residue i, can be obtained from the ratio between the potential $V(i\rightarrow 0)$ produced by residue i at the mutation site (residue 0), in the presence and absence of bulk solvent:[30]

$$\alpha_i = V_w(i\rightarrow 0)/V_v(i\rightarrow 0) \tag{1}$$

where the subscripts v and w refer to either vacuum or bulk water surrounding the system. These ratios measure the screening of the potential $V(i\rightarrow 0)$ by bulk solvent. We shall refer to this scheme as potential-based scaling. The potentials can be obtained from a finite-difference continuum calculation, as described below.

Another protocol is to not use any charge scaling. If the resulting vacuum-induced distortions are too large, one can restrain selected ionized groups to their experimental positions, when known. While restraint energies are likely to be large in the vacuum environment, they should be much smaller after transfer to bulk solvent and contribute little to the overall free energy difference. Their exact contribution to the difference in binding free energies can be calculated with a free energy perturbation approach if desired. This would involve performing two additional steps in the thermodynamic cycle (Figure 1) to reduce the restraints to zero for both M_a and M_b. The two additional free energy contributions would be expected to be small, and cancel out to a large extent.

In what follows we describe steps I−III in more detail.

2.2. Step I: Free Energy Perturbation with Molecular Dynamics. In step I, the macromolecules are modelled microscopically, along with a finite number of explicit water molecules. Selected charges may be scaled and bulk water is not included.

We perform the mutation $M_a\rightarrow M_b$ using thermodynamic perturbation theory.[43,44] In the applications below, we rely mainly on thermodynamic integration schemes, which make use of the first, and sometimes the second, derivatives of the free energy G with respect to a coupling parameter λ; these are calculated as

$$\frac{\partial G}{\partial \lambda}(\lambda) = \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$$

$$\frac{\partial^2 G}{\partial \lambda^2}(\lambda) = -\frac{1}{kT}(\langle U^2\rangle_\lambda - \langle U\rangle_\lambda^2) \tag{2}$$

Here, U is the λ-dependent potential energy function, and the brackets indicate ensemble averages performed at a given λ. In practice, averages over molecular dynamics simulations are used. As the coupling parameter varies from 0 to 1, the energy function $U(\lambda)$ is mapped gradually from that of M_a to that of M_b. A series of simulations at intermediate values of λ provide derivative values that can be numerically integrated[19,45,46] to obtain the desired free energy change.

2.3. Step II: Restoring the Normal Charge Values. If a subset of selected charges have been scaled in the microscopic models of step I, step II removes the scaling. The free energy of this process is obtained as follows.

The system, consisting of a portion of the protein and a finite number of water molecules, is surrounded by vacuum in this step (Figure 1). We require the difference between the free energy changes for "unscaling" M_a and M_b. For the present argument we consider a mutation that adds a single charge q at position 0: $q_0 = 0$ for M_a, $q_0 = q$ for M_b. The free energy difference is then

$$\Delta\Delta G_{M_a\rightarrow M_b}(\{\alpha_i\}\rightarrow 1) = \sum_{i>0} q(1-\alpha_i)V_v(i\rightarrow 0) \tag{3}$$

where $\{\alpha_i\}$ are the scaling factors, which differ from unity for the selected charges.

With potential-based scaling, substituting for α_i from eq 1, we obtain

$$\Delta\Delta G_{M_a\rightarrow M_b}(\{\alpha_i\}\rightarrow 1) = \sum_{i\neq 0,\alpha_i\neq 1} q(V_v(i\rightarrow 0) - V_w(i\rightarrow 0)) \tag{4}$$

Only scaled charges contribute to the sum. As emphasized, the system considered so far is in vacuum; bulk solvent will be introduced in step III. The appearance in eq 4 of $V_w(i\rightarrow 0)$ (the potential produced by i in bulk solvent) is simply due to the definition of α_i. Potential-based scaling has the advantage that each scaled charge contributes a term that cancels exactly its contribution in step III, as shown below (eq 12). The choice of structures used in this step poses a practical problem. In general, the structures of M_a and M_b, in vacuum and bulk water, are expected to be different. We show below that the effect of these conformational changes can be included in the calculation in a simple way.

The magnitude of typical scaling factors can be estimated by considering a spherical system, of radius a, surrounded by a medium of dielectric constant $\epsilon \gg 1$, with q_0 at its center. Indeed, this geometry mimics a typical simulation setup, with the mutated residue at the center of a roughly spherical model surrounded by water. A charge q_i located at a distance $s_i < a$ from the center produces an image charge, $q_i[(1-\epsilon)/(1+\epsilon)]a/s_i$, at a distance a^2/s_i along the q_0-q_i axis.[47] Its potential at the center and its potential-based scaling factor are

$$V_w(i\rightarrow 0) \approx q_i\left(\frac{1}{s_i} - \frac{1}{a}\right) \tag{5}$$

$$\alpha_i \approx 1 - \frac{s_i}{a} \tag{6}$$

where terms of order $1/\epsilon^2$ are neglected. For $s_i \leq a/2$, α_i varies between 1 and 0.5; as s_i approaches a, α_i goes to zero.

Other scaling schemes, or a combination of schemes, can be used, as long as they reduce the effect of the surrounding vacuum on the structure and dynamics in the microscopic simulations of step I to a reasonable level. For example, if there are ion pairs of opposite charges close together, they might be viewed as a unit and be scaled accordingly. In a spherical system, eq 5 shows that a pair of opposite charges $q_i = -q_j$ produces the same total potential at the center, whether the sphere is surrounded by vacuum or water (and regardless of the pair's position):

$$V_w(i\rightarrow 0) + V_w(j\rightarrow 0) = q_i\left(\frac{1}{s_i} - \frac{1}{s_j}\right) = V_v(i\rightarrow 0) + V_v(j\rightarrow 0)$$

$$\alpha_{i,j} = \frac{V_w(i\rightarrow 0) + V_w(j\rightarrow 0)}{V_v(i\rightarrow 0) + V_v(j\rightarrow 0)} = 1 \qquad (7)$$

This result can be generalized to any neutral group of charges: the solvent polarization induced by such a group does not produce a net potential at the center of the sphere. In practice, this implies that the microscopic free energy result of step I is only weakly sensitive to the exact scaling used for pairs of nearby opposite charges, something that is verified in the application below.

2.4. Step III: Electrostatic Transfer Free Energies into Bulk Water Using a Continuum Model. Step III is the transfer of the finite models into bulk water. The transfer free energies are obtained for both M_a and M_b from continuum electrostatics, as follows.

The continuum model views the system (macromolecule + explicit water) as a set of permanent charges q_i embedded in a cavity. We first consider a mutation $M_a \rightarrow M_b$ that adds a single permanent charge, q, at position 0: $q_0 = 0$ for M_a, $q_0 = q$ for M_b. The calculations are readily extended to mutations that add, delete, or modify more than one charge. The total electrostatic free energy of the system is[48]

$$G_s = {}^1/_2 \sum_i q_i V_s(i) \qquad (8)$$

where the subscript s refers to the medium surrounding the system (vacuum or bulk solvent), $V_s(i)$ is the electrostatic potential at position i, and the summation includes position 0. From the linearity of continuum electrostatics, we can write the potential as a sum of the potentials $V_s(j\rightarrow i)$ produced by the q_j at position i: $V_s(i) = \sum_j V_s(j\rightarrow i)$. Making use of the reciprocity relation $q_j V_s(0\rightarrow j) = q_0 V_s(j\rightarrow 0)$ (valid in any inhomogeneous dielectric medium),[48] and separating out terms containing q_0, we obtain

$$G_s = {}^1/_2 q_0 V_s(0\rightarrow 0) + \sum_{i>0} q_0 V_s(i\rightarrow 0) + {}^1/_2 \sum_{i,j>0} q_i V_s(j\rightarrow i) \qquad (9)$$

The first term is a generalized Born term, representing the interaction free energy of q_0 with the polarization it induces in the surrounding medium. The second term contains the interactions between q_0 and all other charges, including the effect of solvent polarization by the other charges when s = bulk water. The last term represents interactions between the charges other than q_0.

We require the difference $\Delta\Delta G_{M_a\rightarrow M_b}(v\rightarrow w)$ between the transfer free energies of M_a and M_b. In general, the structures of M_a and M_b, in vacuum and bulk water, will all be different. Subtracting the vacuum and bulk water free energies leads to terms of the form

$$V_w(i\rightarrow j) - V_v(i\rightarrow j) = V_v(i\rightarrow j)\left(\frac{V_w(i\rightarrow j)}{V_v(i\rightarrow j)} - 1\right) \qquad (10)$$

The term in parentheses is normally close to -1 (since $V_w \ll V_v$) and should be insensitive to the exact molecular conformation. This is confirmed in the application to protein−ligand binding below. In what follows, we assume that it can be calculated using a single representative structure for all four states: M_a and M_b, in water or vacuum (*e.g.*, in the application below, a crystal structure of M_a is used). The multiplying factors $V_v(i\rightarrow j)$ are more sensitive to the molecular conformation. They can in fact be obtained from the molecular dynamics simulations carried out in step I. In the application below, averages are taken over the two endpoints of the microscopic free energy simulation. In this way conformational flexibility is taken into account approximately for the vacuum potentials but is neglected when the ratios of water and vacuum potentials are calculated. With this assumption, the cross-terms involving $i,j \neq 0$ disappear from the free energy differences (see below).

We have assumed for this argument that M_a and M_b differ by a single charge, $q_0 = 0$ for M_a, and $q_0 = q$ for M_b. The difference in binding free energies is

$$\Delta\Delta G_{M_a\rightarrow M_b}(v\rightarrow w) = \Delta G_w - \Delta G_v \qquad (11)$$

where $\Delta G_s = G_s(M_b) - G_s(M_a)$ (eq 9). The $q_i V_s(j\rightarrow i)$ interaction terms cancel in ΔG_s, and the q_0 terms are zero when s = vacuum. We obtain finally

$$\Delta\Delta G_{M_a\rightarrow M_b}(v\rightarrow w) = {}^1/_2 q V_w(0\rightarrow 0) + \sum_{i>0} q(V_w(i\rightarrow 0) - V_v(i\rightarrow 0)) \qquad (12)$$

For a charge that was scaled in step I using potential-based scaling, the free energy contribution of removing the scaling (eq 4) exactly cancels the contribution of that charge in eq 12. Thus, the potential-based scaling of a selected charge, within the continuum model, is formally equivalent to restoring the full solvation of that charge by transferring the microscopic model into bulk water. In effect, potential-based scaling provides a way to "solvate" selected charges *before* carrying out the microscopic free energy perturbation calculation in step I, even though the system is not yet in bulk solvent. A protocol that does not use any charge scaling carries out the solvation entirely *a posteriori*, in step III. Comparison of the two protocols provides a test of the consistency and robustness of the method.

The calculation can be extended to the general case of two macromolecules M_a, M_b, that differ by a set of p charges: *i.e.*, $\{q_i^a, i = 1,2,...,p\}$ are changed into $\{q_i^b, i = 1,2,...,p\}$. We obtain

$$\Delta\Delta G_{M_a\rightarrow M_b}(v\rightarrow w) = {}^1/_2 \sum_{i=1}^{p}\sum_{j=1}^{p} [q_i^b(V_w^b(j\rightarrow i) - V_v^b(j\rightarrow i)) - q_i^a(V_w^a(j\rightarrow i) - V_v^a(j\rightarrow i))] + \sum_{i=1}^{p}\sum_{j>p} (q_i^b - q_i^a)(V_w(j\rightarrow i) - V_v(j\rightarrow i)) \qquad (13)$$

The superscript m = a, b in $V_w^m(j\rightarrow i)$ indicates that the source charge is q_j^m; similarly for $V_v^m(j\rightarrow i)$.

2.5. Example: Application to a Protein−Ligand Binding Problem. In the present application, we compare the complexes of aspartyl-tRNA synthetase with its substrate aspartic acid and

with the substrate analogue asparagine. The first complex plays the role of M_a, the second plays the role of M_b. The mutation, Asp→Asn, inserts a single positive charge on the ligand, *i.e.*, $q = +1$.

The proposed method requires two computations. The first is the precalculation of the electrostatic potentials $V_v(i\rightarrow 0)$, $V_w(i\rightarrow 0)$ produced by the charged groups i at the mutation site, by numerical solution of the Poisson−Boltzmann equation. The second is the microscopic Asp→Asn free energy simulation, using molecular dynamics. The exact molecular structure used in the Poisson−Boltzmann calculations is not critical, following the argument given (eq 10). The ratios $V_w(i\rightarrow 0)/V_v(i\rightarrow 0)$ from the Poisson−Boltzmann computation provide reasonable scaling factors to use in the microscopic free energy simulation.

The free energies for steps II and III depend on which charged groups are scaled, and how. Charges that undergo potential-based scaling in step I do not contribute in steps II and III because the charge unscaling exactly cancels the transfer into bulk solution. Charges that are scaled according to some other scheme do contribute through eqs 4 and 12. Charges that are not scaled at all contribute to step III (eq 12). The present application shows that the results are quite robust with respect to the choice of scaled charges, if it is reasonable, and to the exact scaling protocol.

3. Computational Details

3.1. Description of the Microscopic Model. The microscopic model consists of the active site region of one dimer of the *Escherichia coli* aspartyl-tRNA synthetase enzyme, one ligand molecule (aspartic acid or asparagine) bound in the active site, and several hundred explicit water molecules solvating the active site. Initial protein heavy atom positions were taken from the crystal structure of the enzyme−aspartyl-tRNA complex.[49] The initial position of the aspartic acid ligand was inferred from the crystal structure of the homologous *Thermus thermophilus* enzyme complexed with an aspartyl adenylate ligand[50] by superimposing conserved backbone segments in the active sites from the two structures. Initial hydrogen positions were built from stereochemical considerations.[51] The protonation state of the histidine residues was inferred from inspection of hydrogen bonding in the crystal structure. Two hundred steps of steepest descent minimization were done to relax bad contacts in the crystal structure and further optimize hydrogen positions. An approximately spherical region, of radius 20 Å, centered on the side chain of the ligand, was then selected. It contained 3965 protein atoms (including polar and nonpolar hydrogens) and the ligand. Protein backbone and side chain groups were not subdivided but were either included or excluded as a whole. Care was taken not to split pairs of charged residues forming salt bridges. The final model contained 56 charged residues, with a net charge of −8, not including the ligand.

A sphere of water molecules taken from a water droplet simulation[52] was overlaid on the active site, and molecules that overlapped protein atoms, or were outside a 20 Å sphere, were eliminated. The water sphere was actually offset by 7 Å compared to the protein sphere; *i.e.*, it was further out of the active site, to increase the extent of explicit solvation. No crystal waters were included, due to the moderate resolution of the crystal structure.[49] Several additional water overlays were done during the course of equilibration of the system with molecular dynamics, giving a final total of 384 water molecules, mostly located in or near the deep active site pocket. The model contains 5140 atoms in all; it is shown in Figure 2. The force field model and boundary conditions are detailed in Section 3.5.

3.2. Scaling Selected Protein Charges. Three different scaling protocols were compared in a series of five Asp→Asn free energy calculations. All of these protocols are designed to mimic solvent screening so as to reduce the distortion of the structure during minimization and dynamics in vacuum to a moderate level. They are not required to quantitatively model the full effect of solvent screening on the free energies; following the derivation in the previous section, this is done in steps II and III of the procedure (Figure 1).

There is no obviously optimal scaling protocol, but several features are desirable. First, we limit charge scaling to a subset of ionized protein side chains; no scaling is applied to polar but neutral backbone or side chain groups. Second, we attempt to minimize the number of scaled side chains. Thus charged side chains that are isolated and completely unsolvated in our model are scaled, but pairs of ionized side chains forming salt bridges usually are not. Third, in most cases we choose the number of side chains that are scaled such that the remaining unscaled ones have a total charge of zero. However, two mutation runs were done with a net charge of −5 for the unscaled residues, providing a stringent test of the *a posteriori* continuum treatment of the vacuum→bulk water transfer step. Finally, we preferentially choose for scaling the side chains that have the largest potential-based scaling factors, *i.e.*, the side chains that are most strongly screened by bulk solvent, and are presumably most affected by the absence of bulk solvent in step I.

With these criteria, we chose for scaling an initial set of 19 charged side chains, largely unsolvated by explicit waters in the microscopic model (Table 1). The remaining unscaled residues had a net charge of −5, not including the ligand. Two mutation runs were performed with only the 19 side chains scaled. This scaling scheme will be referred to as protocol A.

Two alternative sets of seven additional residues to scale were identified, each set having one positive and six negative residues and leaving a net charge of zero for the unscaled side chains. The additional residues were chosen because they had large potential-based scaling factors, despite being partly solvated by explicit waters. Each set provides a reasonable choice. Runs were performed with either set scaled. These "neutral" scaling schemes will be referred to as protocols B and C, respectively (Table 1).

Potential-based scaling factors were used in most cases. However in the first two runs (protocol A), slightly smaller scaling factors were used. This happened inadvertently, due to the presence of small cavities in the protein interior, which were treated as solvent pockets by the finite-difference Poisson programs (see below); *i.e.*, the electrostatic potentials calculated initially contained artefacts, and the scaling factors were systematically smaller than the true potential-based ones. In subsequent runs (protocols B and C), these continuum solvent pockets were eliminated. Thus the first two runs provide an additional test for the *a posteriori* continuum corrections in steps II and III, which restore the full charges and transfer the system into bulk water. The three scaling protocols are summarized in Table 1.

3.3. Continuum Poisson−Boltzmann Calculations. Continuum calculations are required to obtain the potentials produced on the ligand by each charged group in the system. The entire enzyme dimer, the ligand, and the explicit water molecules from the microscopic model make up a dielectric cavity, immersed either in a surrounding vacuum or in bulk water. It is important to note that the entire dimer must be included in the calculation, to correctly exclude water from all regions that are occupied by protein, since the active site region used in the microscopic simulations alone would not accomplish this. Calculations are done successively with each charged side

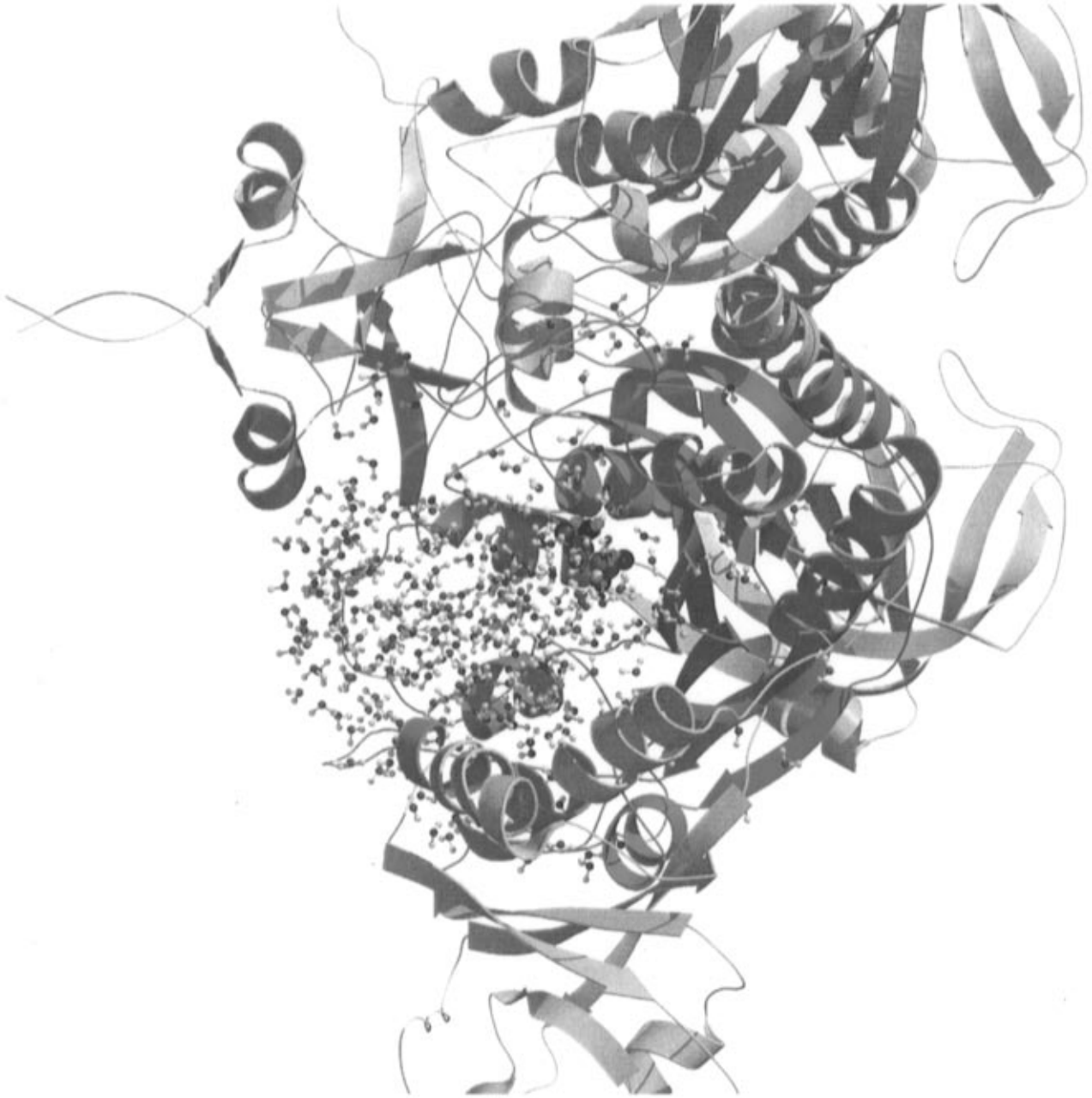

Figure 2. Schematic view (C_α trace) of the aspartyl-tRNA synthetase dimer. The upper part of the dimer is left out. A roughly spherical protein region (green residues) and 384 explicit waters (ball-and-stick models) were treated by molecular dynamics; protein atoms outside the spherical region were kept fixed. The ligand is shown in CPK representation, near the center of the spherical region. Most of the waters are located in or near the active site cleft; however, some of the waters make up a thin shell extending partway around the surface of the protein (*e.g.*, there are isolated waters in surface pockets that can be seen on the right-hand side of the simulation region).

TABLE 1: Charge Scaling Protocols Used

protocol	potential-based[a]	residues scaled
A	no	Lys167, Asp207, Arg208, Arg245, Glu246, Glu249, Glu274, Glu276, Asp285, Lys399, Lys400, Asp432, Glu437, Asp438, Glu467, Lys455, Glu509, Lys513, Arg549
B	yes	same as protocol A, plus Glu171, Asp282, Arg287, Asp404, Asp439, Asp546, Glu578II[b]
C	yes	same as protocol A, plus Glu219, Arg287, Asp404, Asp439, Asp536, Asp546, Glu578II

[a] Scaling factors given by eq 1. [b] The "II" indicates a residue from the second enzyme monomer.

chain *i* as the sole source charge. The resulting potential on each ligand atom is calculated, and averaged over the ligand. The scaling factor α_i is the ratio of the results in bulk water and in vacuum (eq 1). The charged side chains included in this calculation are the ones included in the microscopic free energy simulations in step I. Thus residues outside the 20 Å microscopic sphere, while they help to make up the cavity, are not included as source charges.

The enzyme crystal structure was employed,[49] along with the 384 explicit waters that were added (see above). The water positions were obtained from a short molecular dynamics run with a fixed protein. No crystal waters were used, due to the moderate resolution of the crystal structure. The role of the explicit water in the present calculation is to exclude bulk solvent from the microscopic simulation region. The outer boundary of the simulation sphere is 20 Å from the ligand side chain, and so the exact water positions, and the fine details of the dielectric boundary between the cavity and the outer medium, are not critical.

The Poisson equation was solved using standard finite-difference methods.[53−55] An initial large cubic grid is overlaid on the system, and a series of "focussing" calculations are done with successively smaller and finer grids. The final grid spacings were about 1.3 Å. This is sufficient given that the source charge is always at least 10−20 Å away from the ligand. Calculations were done with several values for the ionic strength, and with several slightly different definitions of the cavity−high-dielectric boundary. All calculations were done with both the Delphi[54] and the UHBD[56] programs. As mentioned above, these programs found three or four small cavities in the protein interior, which would have been treated as high dielectric constant pockets. To avoid this, the pockets were explicitly filled with uncharged dummy atoms, which have a dielectric constant of unity. The potential values finally used were obtained with zero ionic strength, and a 0 Å probe radius for the construction of the solvent−cavity boundary. The results were of the same magnitude as values obtained with 0.1 M ionic strength (concentration of monovalent ions) and a 1.4 Å probe radius.

3.4. Comparison with a Spherical Reaction Field Treatment of Solvent. The continuum contributions in steps II and III are also calculated, for comparison, using a cavity that contains only the microscopic model as opposed to the complete enzyme dimer. In this case, the microscopic model is immersed in water, but the low-dielectric region formed by the rest of the enzyme dimer is not taken into account. Since the microscopic model is approximately spherical, these conditions approximate the assumptions that are made in a spherical reaction field treatment of the solvent.[47] If the results are in good agreement with the complete, inhomogeneous, dimer model, then the simpler spherical approximation could be used in future applications. This would have the advantage that the solvent reaction field could be calculated "on the fly" during the molecular dynamics simulations,[24−26] so that there would be no need for *a priori* charge scaling or for an *a posteriori* vacuum→water step.

3.5. Force Field and Molecular Dynamics Simulations. Atomic charges, van der Waals, and stereochemical force field parameters for protein and ligand were taken from the CHARMM22 all-atom force field.[57] Water interactions were described by a modified TIP3P model.[58,59] The dielectric constant of this model is 82 ± 4, based on simulations in a spherical geometry similar to the one used here.[52] Electrostatic interactions were treated essentially exactly, without any form of truncation, in an efficient way by using a multipole approximation ("extended electrostatics") for groups more than 12 Å apart.[60,29] Van der Waals interactions were truncated at an interatom separation of 13 Å.

Protein atoms more than 15 Å from the center were harmonically restrained to their initial positions, with force constants determined from the average observed crystallographic *B*-factor of the residue to which each atom belongs, and are multiplied by a scaling function that increases smoothly from zero to one as the atomic distance from the sphere center increases from 15 to 20 Å.[27] Bond lengths to hydrogen atoms, and the internal geometry of the water molecules, were constrained with the SHAKE algorithm to standard values.[61]

Water molecules were confined to a spherical, 20 Å radius, region by the stochastic boundary method.[27,28] The mean field boundary potential is based on the average interactions in bulk water at equilibrium, and so it does not account for the solvent reaction field induced by charged protein side chains located within the simulation sphere. Thus from the point of view of electrostatics, the system is immersed in a medium with no dielectric response, namely a vacuum. This is corrected for with the continuum model (steps II and III in the thermodynamic cycle, Figure 1).

Protein and water heavy atoms located more than 15 Å from the center of the simulation sphere experience frictional and random forces that mimic the flow of energy between the system and a surrounding thermal bath at 293 K.[62] Equations of motion are integrated with the Verlet algorithm, with a 1 fs time step, using the CHARMM program.[60]

The model is shown schematically in Figure 2.

3.6. Free Energy Calculations. The free energy calculations model the alchemical transformation of the native aspartic acid substrate into an asparagine molecule in the protein−ligand complex. A single ligand molecule is introduced, but with two side chains, one of each type. The mutation is performed by "growing in" one side chain while "growing out" the other. The approach used here is based on a modified version of the BLOCK module in the CHARMM program.[63,64]

Specifically, the system is divided into three "blocks": the aspartate side chain atoms (block 2), the asparagine side chain atoms (block 3), and the rest of the system (block 1); the backbone atoms of the ligand are included in block 1. The potential energy function has the form

$$U(\lambda) = U^{b} + U_{11}^{nb} + U_{22}^{vdw} + U_{33}^{vdw} + \lambda(U_{12}^{nb} + U_{22}^{elec}) + (1-\lambda)(U_{13}^{nb} + U_{33}^{elec}) + U_{res}(\lambda) \quad (14)$$

where U^{b} represents all the bonded interactions in the system, including those involving the ligand; U_{11}^{nb} represents the nonbonded interactions within block 1; U_{ii}^{vdw} and U_{ii}^{elec} represent the van der Waals and electrostatic interactions within block i (= 2 or 3); and U_{1i}^{nb} represents the nonbonded interactions between blocks 1 and i. $U_{res}(\lambda)$ is a restraint term that gradually forces the ligand to move from the initial aspartic acid binding position into a stable asparagine binding site, identified in separate simulations.[41] This term is discussed below. Bonded interactions between blocks 1 and 2, and between blocks 1 and 3, are not scaled, in contrast to earlier applications of the BLOCK module in CHARMM, *e.g.*, refs 63 and 65. This lack of scaling avoids endpoint problems associated with covalent energy terms (*e.g.*, bond terms), which can arise when groups disappear or appear. By so doing, certain contributions to the free energy are omitted from individual portions of the thermodynamic cycle (*i.e.*, Asp→Asn in the protein or in solution). However these contributions are expected to be small in the present application and are known to cancel in the full cycle.[66]

The "dual topology" method used here has the advantage that the conformations of the reactant (Asp) and product (Asn) side chains are not constrained to coincide, but can adjust separately to the environment. The linear λ scaling is known to lead to large free energy derivatives near the mutation endpoints, due to repulsive van der Waals interactions.[19] For simple systems, such as van der Waals liquids, the free energy derivative was shown theoretically[67,68] and numerically[68,69] to go to infinity as $\lambda^{-1/4}$ when $\lambda \rightarrow 0$. For more complex systems like proteins, this functional form appears to be approximately verified, and can serve to guide a fitting process that avoids the end point problem.

Simulations were run with λ = 0.02, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 0.98, and the first and second derivatives $\partial G/\partial\lambda$ and $\partial^2 G/\partial\lambda^2$ were obtained from eq 2. Numerical integration was done using a standard trapezoidal method. Endpoints (*e.g.*, $\lambda = 0 \rightarrow 0.02$) were treated either by polynomial extrapolation of the free energy based on its first two derivatives (van der Waals and electrostatic terms) or by fitting a function

$A\lambda^{-1/4}$ to the first two $\partial G/\partial\lambda$ values (van der Waals terms). An upper bound for the numerical error associated with trapezoidal integration is obtained from the theory of Newton−Cotes integration methods.[4,70]

Simulations at each λ included 10 ps of equilibration and 30 ps of data collection. In our runs, the relaxation times of $\partial U/\partial\lambda$ were found to be much smaller than 10 ps, with the exception of electrostatic terms in a few windows, which relaxed in ∼10 ps. Thus 10 ps should provide adequate equilibration after each change in λ.[71,72] The rapid convergence is likely to result in part from the lack of bulk solvent. Random statistical errors[73,42] were sufficiently small with this window length. Each run required a week of CPU time on a Cray C90 supercomputer.

Five mutation runs were performed, three in the Asp→Asn direction and two in the Asn→Asp direction, with various charge scaling protocols as outlined above and in Table 1. Runs in opposite directions were prepared and equilibrated independently of each other.

3.7. Starting Structures and Ligand Restraints. The native Asp−protein complex is very stable in simulations, with a single well-defined ligand position. The Asn ligand, on the other hand, can occupy several slightly different positions in the active site. After five molecular dynamics simulations of the Asn−protein complex, with different ligand starting geometries, one of the more stable positions was selected as a target structure for the free energy runs. This structure was 1.1 Å from the Asp position. A simple linear pathway was constructed, interpolating between the Asp binding position and the chosen Asn binding position. During the free energy calculations, a harmonic umbrella restraint term was applied to the ligand backbone heavy atoms. In some runs (1, 2, and 5), the ligand side chain was restrained as well. The target positions for these umbrella restraints varied as a function of λ along the interpolation pathway. At each position, the restraint potential biases the simulation and the resulting ensemble averages, such as $\langle\partial U/\partial\lambda\rangle$. To obtain the unbiased ensemble averages, we use the standard relation[74]

$$\langle A\rangle_{\text{free}} = \langle Ae^{U/kT}\rangle_{\text{res}}/\langle e^{U/kT}\rangle_{\text{res}} \qquad (15)$$

where the subscripts "free" and "res" refer to the unbiased and "restrained" averages, respectively. In the application below, the "unbiasing" via eq 15 has a small effect on the overall free energy changes.

4. Results

4.1. Potential-Based Scaling Factors and Choice of Scaled Residues. The screening of each charged side chain *i* by bulk solvent is characterized by the screening factor $1/\alpha_i$, the inverse of the potential-based scaling factor. These screening factors are shown in Figure 3 for all side chains as a function of distance from the ligand, which is at the center of the approximately spherical microscopic region (Figure 2). They vary between 1 and 60, increasing rapidly for distances greater than 15 Å, though not as rapidly, on average, as a simple spherical model predicts. In the Poisson−Boltzmann calculations, the dielectric cavity is formed by the entire enzyme dimer; on the other hand the cavity's outer boundary is rough, so that the bulk solvent region makes incursions within the nominal 20 Å radius of the microscopic region. This leads to screening behavior for the system that is more complicated than that for a dielectric sphere.

The total protein charge within a given distance from the ligand is seen to be positive for distances up to 8 Å, then close to zero up to 14 Å, and finally to decrease rapidly to −8e. The

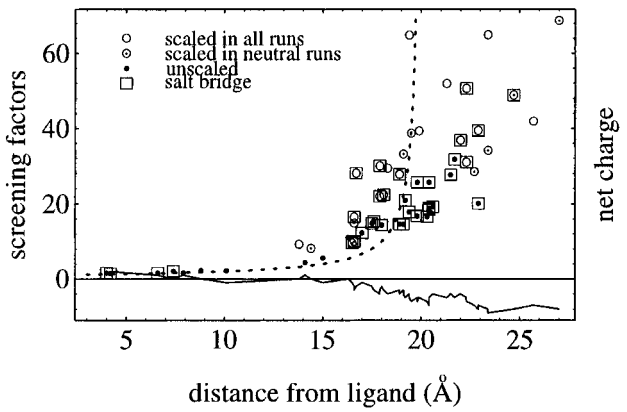

Figure 3. Bulk solvent screening factors $1/\alpha_i = V_v(i\rightarrow 0)/V_w(i\rightarrow 0)$ from continuum model. Residues involved in salt bridge pairs are highlighted by boxes. Dashed line: screening factors predicted by spherical cavity model. Solid line: total net protein charge within the given distance from the ligand (excluding the ligand charge). "Neutral" runs use protocols B or C (Table 1).

distant negative charges are highly screened, not only by the bulk solvent but also by explicit solvent in the microscopic region.

Residues whose charges were scaled in the free energy calculations are indicated in Figure 3, and listed in Table 1, which shows the choices that were made in the different scaling protocols A, B, and C. We emphasize again that for all charged residues, the effect of transfer into bulk solvent is calculated with the same continuum approach. The use of different scaling protocols does not affect the initial and final states that are being compared; it merely represents slightly different pathways between the two states. The *a priori* scaling of selected charges modifies only the structures and energies sampled during the molecular dynamics and hence the results of the microscopic free energy calculations. However, this is corrected for, in principle, in steps II and III of our cycle (Figure 1). By comparing three different scaling protocols, we show that the correction works in practice and is robust with respect to the choice of screened charges.

Most of the residues with large screening factors were scaled in all five free energy runs. Some others are involved in salt bridges, so that the free energy results are not expected to be sensitive to their scaling (see Theory). This is confirmed by comparing runs where particular salt bridges were scaled, to runs where they were not scaled (*e*.g., residues in Arg141−Asp536 and Arg174−Glu219 were scaled only in run 5, and residues in Asp285−Arg287 were scaled in runs 1−4 but not in run 5). A few residues not involved in salt bridges, but with large screening factors, were only scaled in one or two of the runs (*e.g.*, Arg546, see Table 1). It is found in the calculations that whether or not they are scaled has a small effect on the final free energies. Thus, the *a priori* scaling during the molecular dynamics calculations and the *a posteriori* bulk solvent correction give very similar results for the free energy contributions of these residues. Arg546, the outermost charged residue at a distance of 27 Å (Figure 3), has its position restrained in the molecular dynamics simulations, so that scaling has little effect on its position and dynamics. The six unscaled residues not involved in salt bridges (Asp175, Arg217, Arg225, Asp475, Glu482, Arg537) are all well solvated by explicit water. Two of them can also be grouped in a neutral pair separated by 4 Å, although they do not form a salt bridge.

4.2. Microscopic Free Energy Simulations: Statistics. There are many possible sources of error in free energy calculations. However with care, most can be analyzed and

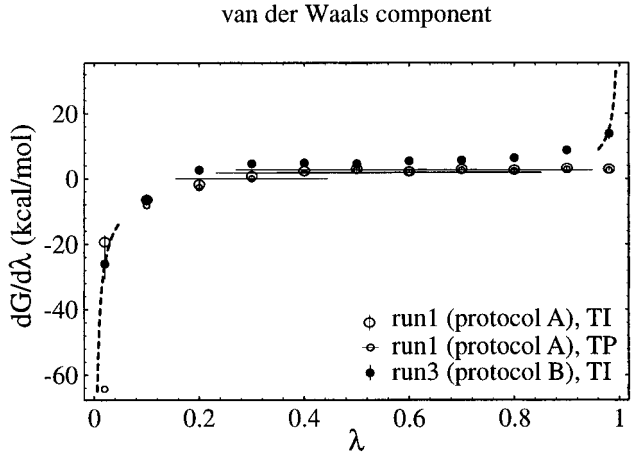

Figure 4. Van der Waals component of $\partial G/\partial\lambda$ for two typical runs (1 and 3), which used scaling protocols A and B, respectively. Thermodynamic integration (TI) and thermodynamic perturbation (TP) results (from exponential formula; shown for run 1 only). Error bars are shown as vertical lines (TI data only, see text). For most data points the error bars are too small to be seen. The TI and TP points often overlap. Horizontal bars through the TP points indicate the acceptable step size along λ for the exponential formula, predicted from the standard deviation of $\partial G/\partial\lambda$ (see text). Horizontal bars are omitted for the data points closest to the end points, where the acceptable step size cannot be predicted in this way. Dashed curves near the end points represent an extrapolation of the derivative to the end points with the theoretical form $A\lambda^{-1/4}$ or $A(1-\lambda)^{-1/4}$ (see text).

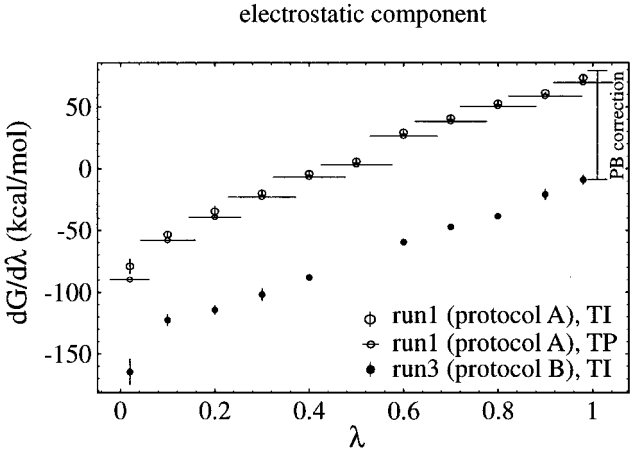

Figure 5. Same as Figure 4, electrostatic component. For many data points the error bars are too small to be seen, and the TI and TP points often overlap. The difference between the continuum corrections for protocols A and B is shown as a vertical line on the right ("PB correction").

controlled in a quantitative way. The main exception is the systematic error introduced by the choice of potential function, which cannot be easily quantified.

In this work, the microscopic free energies are obtained by numerically integrating the first, and in some cases the second, derivatives of the free energy, obtained at successive λ values. The main contributors to the first derivatives are the van der Waals, and especially the electrostatic, interactions between the hybrid side chain and the rest of the system. A third contribution comes from interactions within the hybrid side chain, but this contribution is almost identical for the Asp→Asn mutations in the protein and in solution (to within <0.3 kcal/mol, *i.e.*, within statistical uncertainty), and so it cancels almost completely in the overall thermodynamic cycle used to study ligand binding.[41] Because these "intra-side chain" contributions have essentially no effect on the structure and dynamics of the system and have no bearing on solvation or binding, we do not discuss them further.

The van der Waals and electrostatic components of the free energy derivatives are shown in Figures 4 and 5 for two typical free energy runs (runs 1 and 3; Table 4), which used protocols A and B, respectively. In protocol A, the unscaled residues have a net charge of −5e, while in protocol B the unscaled residues have a zero net charge. The van der Waals component near the end points, in three out of four cases, has the typical shape associated with creating or deleting atoms in a condensed environment. The large end point tails resemble those commonly seen, and predicted theoretically, for van der Waals liquids,[67,68] which go to infinity as $\lambda^{-1/4}$. However, at the Asn end of run 1 (see Figure 4), the Asp side chain moiety of the ligand has more limited χ_1 and χ_2 torsional excursions than in the other runs. This gives a small van der Waals free energy derivative contribution, and the end point tail is not observed. This is due to the positional restraints applied to the ligand side chain in this run (see Methods), which result in a greater overlap between the Asp and Asn side chain moieties. Extrapolation of the derivative to $\lambda = 0$ and 1 was done either assuming a constant second derivative (linear extrapolation), or assuming the $\lambda^{-1/4}$ (or $(1-\lambda)^{-1/4}$) form. The free energies reported use the average result from the two extrapolation methods, whose results differ typically by about 0.5−1 kcal/mol.

The electrostatic free energy derivative increases smoothly with λ (Figure 5), in a roughly linear fashion, corresponding to a slowly varying second derivative. In fact, the second derivative is about two times larger at the Asp end point than at the Asn end point. This indicates that the dielectric susceptibility of the system, in response to a changing charge on the ligand, decreases by a factor of 2 in going from Asp to Asn. The fluctuations of the electrostatic potential on the ligand side chain, which measure this susceptibility, have roughly the same Gaussian shape throughout the various runs but are broadened more at the Asp end point. This nonlinearity of the free energy derivative represents a departure from a linear response behavior, which would correspond to a constant second derivative.

Also shown in Figures 4 and 5 are the local free energy slopes calculated from the thermodynamic perturbation method, *i.e.*, from the so-called exponential formula:[19]

$$\frac{\partial G}{\partial\lambda}(\lambda_i) \approx \frac{G(\lambda_{i+1}) - G(\lambda_i)}{\lambda_{i+1} - \lambda_i}$$

$$G(\lambda_{i+1}) - G(\lambda_i) = -kT \ln\langle\exp\{-[U(\lambda_{i+1}) - U(\lambda_i)]/kT\}\rangle_i \quad (16)$$

We see that the exponential formula becomes unreliable for the van der Waals component near the $\lambda = 0$ end point. This can be traced to a very short, abrupt negative tail in the probability distribution of $U(\lambda_{i+1}) - U(\lambda_i)$, which does not provide sufficient overlap with the exponential in eq 16 (see *e.g.* refs 10 and 75). On the other hand, the free energy derivative is only weakly sensitive to this tail and numerically much more reliable. Horizontal bars through the data points in Figures 4 and 5 measure kT/σ_i, where σ_i is the standard deviation of $\partial U/\partial\lambda$ in window i. This represents the approximate step size that is permissible for λ in the exponential formula.[10] Bars are omitted near the van der Waals end points, because there $\partial U/\partial\lambda$ deviates strongly from a Gaussian distribution, and the simple relation between σ_i and step size then breaks down.

Vertical bars through the data points show the small uncertainty due to random statistical noise, calculated as the

standard deviation of 10 ps block averages (over each 30 ps simulation window). Statistical errors estimated from the autocorrelation function of $\partial U/\partial\lambda$ 76 were systematically smaller and are not shown. Time-lag hysteresis effects should also be small, since the quantity $\partial U/\partial\lambda$ appears to relax in just a few picoseconds in most windows from inspection of its time autocorrelation function (not shown). Furthermore, free energy runs in the Asp→Asn and Asn→Asp directions were prepared and equilibrated independently of each other. To measure time-lag hysteresis directly, we also performed runs starting from one of the end points, extending one third of the way to the other end point, and then back again to the starting point (*e.g.*, $\lambda = 0 \rightarrow 0.35 \rightarrow 0$). One of these runs was performed from each end point. These runs exhibited a free energy hysteresis of 0.7 and 1.9 kcal/mol, respectively, which can probably be accounted for by small differences in the end point structures, as well as random statistical error.

The derivative $\partial U/\partial\lambda$ was integrated with a trapezoidal method, except at the end points of each run. An upper bound for the error in the numerical integral can be obtained if an upper bound for the third derivative $\partial^3 G/\partial\lambda^3$ is available.[45,70] The third derivative can be estimated from the simulations, either as a finite-difference of second derivative values, or directly from the third-order term in the Taylor expansion of the perturbation free energy.[77] Both methods are very imprecise for the present simulation length and only give third-derivative values at the eleven λ values actually simulated. Nevertheless, as the third-derivative estimate is much larger (5−10 times larger) at the Asn end point than in any other window, it is likely that this value represents a reasonable estimate of the upper bound. We then find that over the interval $\lambda = 0.05-0.95$, an upper bound for the integration error is 1.6 kcal/mol. The actual error may be much smaller than this. For example, if we use the next largest value of the third derivative, we obtain an error estimate of only 0.3 kcal/mol.

Umbrella corrections are reported in Table 4. They range from −2.4 to 2.8 kcal/mol. The sign of the correction is the same for runs in a given direction, *i.e.*, the correction to ΔG(Asp→Asn) is negative for runs in the Asp→Asn direction, and positive for runs in the opposite direction. This presumably reflects a small hysteresis in the simulations.

4.3. Microscopic Free Energies. The results for the five free energy runs are collected in Table 4. The microscopic free energy to mutate Asp to Asn in the complex varies between −4.3 and 78.8 kcal/mol, depending mainly on the scaling protocol, although much smaller differences are also seen due to variation in other aspects of the individual runs (see the column headed "total" under "microscopic step I" in Table 4 and compare the forward and backward runs with the same protocol). We show in the next section that the very large differences in the microscopic contributions in protocols A, B, and C are almost exactly compensated by differences in the continuum contributions.

The remaining differences between the five runs can be accounted for by sampling effects. Indeed, great care must be taken in free energy calculations to ensure that sufficient and consistent sampling is accomplished, particularly of the side chain torsional degrees of freedom near the mutation site when there are several alternate conformations that are separated by significant energy barriers.[78−82] In the course of the 2.2 ns of molecular dynamics carried out in our free energy simulations, we have monitored all the conformational changes occurring in the vicinity of the Asp binding site. While such changes do occur, most of them are sampled a fairly large number of times. The active site structure is stable overall; that is, all the initial and final end point structures (at either the Asp or the Asn end of the mutation) are in reasonable agreement, falling within the range sampled in separate, equilibrium, 200 ps simulations of the native and mutant complexes.[41] Nevertheless, there are structural and energetic differences as well as uncertainty surrounding the amount of sampling of particular degrees of freedom. For example in run 1, the main conformational changes taking place in the vicinity of the ligand involve residues Lys198 and Gln195. In the Asp complex, they both interact directly with the Asp ligand, hydrogen bonding to the side chain carboxylate and the backbone ammonium groups, respectively. In the Asn complex, they have swung away from the ligand, which itself has moved 1.1 Å away. These conformational changes are also seen in run 2. In run 3, however, these residues occupy slighly different positions at the Asp end point. As noted, all these runs were prepared largely independently of each other. Gln192, Pro194, Gln231, Ser487, and Leu531 also undergo a limited number of conformational changes in some of the runs and occupy slightly different positions in some of the end point structures. None of these five residues interacts directly with the ligand, however.

More detailed analysis of the structures sampled in the simulations will be presented elsewhere.[41]

4.4. Continuum Contributions to the Free Energy Difference ΔG(Asp→Asn). As shown in Theory, the actual calculation of the bulk solvent corrections are slightly different for different sets of residues, depending on whether, and how, they are scaled in the molecular dynamics runs. For residues that undergo potential-based scaling, there is no need to calculate the continuum correction for transfer to bulk solvent: bulk solvation is taken into account automatically by the charge scaling. For residues that are scaled with different scaling factors (*i.e.*, not according to eq 1), the continuum correction must be calculated, but is normally quite small. For residues that are not scaled, we must in principle calculate the continuum contribution explicitly. However in practice, we omit pairs of residues forming salt bridges, on the grounds that when treated as a group, they experience very little solvent screening. In a perfectly spherical dielectric cavity, we saw that the "pairwise" screening factor would be exactly unity (eq 7). In the present system, the pairwise screening factors are in fact close to 1 (not shown). Furthermore, such neutral pairs, which are mostly more than 15 Å from the ligand, only make a small contribution to the Asp→Asn free energy difference regardless of the exact scaling protocol. There are just two salt bridge pairs less than 15 Å from the ligand: the tightly associated pairs Arg489−Glu235 and Lys198−Asp233, 2−7 Å from the ligand. These pairs are well within the portion of the model solvated by explicit water. Their "pairwise" scaling factors are calculated to be 1.01 and 1.00, respectively, so that screening of their potential by bulk water is indeed completely negligible.

Table 2 shows both the potential-based screening factor $1/\alpha_i$ and the unscreened interaction energy $qV_v(i\rightarrow 0)$ with the aspartic acid ligand (eq 12), for each charged residue *i*. As mentioned, the same structure is used in all the Poisson−Boltzmann calculations to obtain the screening factors. In contrast, the interaction energies are averaged over conformations sampled in the molecular dynamics simulations of the Asp−synthetase and Asn−synthetase complexes. These simulations represent the two end points of the alchemical free energy calculations. To assess structural and energetic relaxation in response to the mutation, the interaction energies are also averaged separately over the Asp and Asn simulations, and the difference is taken. The Asp side chain moiety is present in the simulation of the

TABLE 2: Scaling Factors and Ligand Interaction Energies for all Charged Residues[f]

residue *i*	distance[a]	$1/\alpha_i$[b]	$qV_v(i\rightarrow 0)$[c]	$q\delta V$[d]	residue *i*	distance[a]	$1/\alpha_i$[b]	$qV_v(i\rightarrow 0)$[c]	$q\delta V$[d]
Arg141	19.1	14.5	−16.6		*Asp536*	16.6	9.9	19.7	
Arg149	19.4	17.8	−17.0		Asp159	17.5	14.7	18.5	
Glu161	20.4	18.4	15.4		Lys214	18.0	14.3	−17.2	
Arg174	19.8	25.6	−15.7		*Glu219*	16.5	9.5	21.7	
Lys198	4.2	1.4	−74.9	−52.0	Asp233	6.6	1.6	51.5	15.2
Asp207	17.9	30.0	18.6		**Arg208**	17.9	21.9	−18.7	
Asp218	19.2	20.9	16.2		Arg1811II[e]	21.5	27.6	−14.9	
Asp220	21.7	31.7	14.8		Arg222	20.4	25.6	−15.7	
Asp224	17.6	15.3	19.0		Lys283	17.0	12.3	−20.3	2.3
Glu228	19.8	16.7	15.6		Lys556	22.9	20.0	−14.0	
Glu235	7.4	2.0	55.7	17.3	Arg489	4.0	1.6	−105.4	−28.3
Glu276	22.3	30.9	15.9		**Lys399**	22.0	36.8	−15.7	
Asp285	22.9	39.4	14.3		*Arg287*	24.7	48.9	−13.5	
Lys400	18.9	27.8	−18.2	2.2	**Glu437**	18.1	22.3	20.4	−2.2
Asp432	16.6	16.5	21.5	−2.0	**Lys455**	16.7	28.1	−20.8	
Glu509	22.3	50.6	15.1		**Lys513**	16.6	10.1	−19.2	
Glu161II[e]	18.9	14.6	17.3		Lys2141II[e]	20.4	18.8	−15.5	
Glu228II[e]	20.3	16.5	15.9		Lys556II[e]	20.6	19.2	−15.6	
Lys167	19.9	39.3	−16.0		*Glu171*	16.6	14.9	21.7	2.4
Asp175	15.0	5.6	20.7		Arg217	7.9	1.6	−36.4	
Arg225	16.3	11.1	−20.0		**Arg245**	13.8	9.2	−26.4	2.3
Glu246	19.4	64.8	16.9		**Glu249**	18.3	29.4	18.5	
Glu274	23.4	64.9	14.7		*Asp282*	14.4	8.1	23.4	−4.0
Asp404	22.7	28.5	15.4		**Asp438**	21.3	51.9	15.5	
Asp439	19.1	33.1	17.6		**Glu467**	19.5	38.7	17.9	
Asp475	8.8	2.3	41.7	−12.0	Glu482	10.1	2.2	34.4	−6.3
Arg537	14.1	4.4	−27.5		*Asp546*	27.0	68.7	12.2	
Arg549	25.7	41.9	−12.4		*Glu578II*[e]	23.4	34.1	14.2	

[a] Distance from ligand (Å). [b] Potential-based screening factor. [c] Interaction energy with Asp ligand, averaged over Asp− and Asn−synthetase simulations. [d] $q\delta V$ = difference between interaction energies in Asp− and Asn−synthetase complexes. [e] The "II" indicates a residue from the second enzyme monomer. [f] Pairs of residues forming salt bridges are grouped in the upper part of the Table (above empty line). Residues scaled in all three protocols (A−C, Table 1) are in bold. Residues scaled in one or both of the "neutral" protocols (B, C) are in italics. Other residues are never scaled.

Asn end point (with a weight of $\lambda = 0.02$) and is used to obtain the interaction energy at the Asn end point.

While the interactions extend over a long range, only a few residues exhibit large differences in their interaction energies with the ligand side chain, between the initial (Asp) and final (Asn) states of the system. The interaction energy differences $q\delta V$ (only reported when they are larger than 2 kcal/mol) are dominated by just six residues—especially Lys198 and Arg489, which hydrogen bond directly to the ligand side chain, and strongly stabilize the Asp ligand. Asp233 and Glu235, which form salt bridges to Lys198 and Arg489, make smaller contributions of opposite sign. The next largest contributions come from Asp475 and Glu482. The interaction energies change mainly because the ligand position changes by 1.1 Å in going from Asp to Asn. In addition, Lys198 occupies different positions in the two complexes. The contribution from these six residues to the free energy is calculated entirely from the microscopic simulations, which explicitly model the structural relaxation of protein and ligand in response to the mutation.

Table 3 shows the continuum contributions to ΔG(Asp→Asn) from different sets of residues, in the three scaling protocols used. Physically, these free energy contributions arise from the interaction of solvent polarization, induced by each set of residues, with the negative charge on the Asp side chain. The Asp ligand itself contributes a self-energy of 9.5 kcal/mol, favoring the Asp−synthetase complex over the Asn−synthetase one. For a perfectly spherical 20 Å cavity we would have obtained 8.3 kcal/mol. Protocol A used scaling factors for 19 residues; these differed slightly from the potential-based ones, due to artefacts in the initial Poisson−Boltzmann calculations (the presence of several cavities in the structure, which were inadvertently treated as bulk solvent pockets; see Computational Details). Removing the scaling and transferring the system to bulk solvent gives a contribution of 2.6 kcal/mol from these residues. Seven additional residues are to be scaled in protocol B; solvating them in protocol A gives a free energy of 85.3 kcal/mol (favoring Asp). Finally, the six remaining unscaled residues, not involved in salt bridges, contribute 4.5 kcal/mol.

In protocol B, potential-based scaling is used throughout, and seven additional residues have been scaled, leaving a neutral set of residues unscaled. The Asp ligand contributes the same

TABLE 3: Continuum Contributions to ΔG(Asp→Asn) (kcal/mol) in the Three Scaling Protocols Used

protocol	free energy contribution	residue set
A	9.5[a]	Asp ligand[b]
	2.6	19 "misscaled" residues[c]
	4.5	6 unscaled residues, globally neutral, not in salt bridges (Asp175, Arg217, Arg225, Asp475, Glu482, Arg537)
	85.3	Glu171, Asp282, Arg287, Asp404, Asp439, Asp546, Glu578b
	101.9	total
B	9.5	Asp ligand
	4.5	6 unscaled residues, globally neutral, not in salt bridges (Asp175, Arg217, Arg225, Asp475, Glu482, Arg537)
	14.0	total
C	9.5	Asp ligand
	4.5	6 unscaled residues, globally neutral, not in salt bridges (Asp175, Arg217, Arg225, Asp475, Glu482, Arg537)
	10.5	Arg174, Arg141, Glu171, Asp282
	24.5	total

[a] The positive sign indicates stabilization of the Asp−synthetase complex, with respect to the Asn−synthetase complex. [b] Generalized Born self-energy (the first term in eq 12). [c] The scaling factors were slightly different from the potential-based ones (see text).

TABLE 4: Asp→Asn Free Energy Results (kcal/mol) for the Five Runs

			microscopic step I				continuum steps II and III	
run	protocol	direction of run[a]	elec[b]	vdW	umbrella	total	total[c]	grand total[d]
1	A	Asn→Asp	−7.4	0.4	2.7	−4.3	101.9	97.6
2	A	Asp→Asn	−2.1	−0.3	−1.3	−3.7	101.9	98.2
3	B	Asn→Asp	76.6	−2.4	2.8	77.0	14.0	91.0
4	B	Asp→Asn	83.5	−3.7	−1.0	78.8	14.0	92.8
5	C	Asp→Asn	75.1	0.2	−2.4	72.9	24.5	97.4
								average: 95.4 ± 2.9

[a] Runs in opposite directions were prepared and equilibrated independently of each other. [b] Not including intraresidue contribution, which is essentially constant over all runs (and in the Asp→Asn mutation in solution). [c] See details in Table 3. [d] Free energy change for the Asp→Asn mutation in enzyme active site.

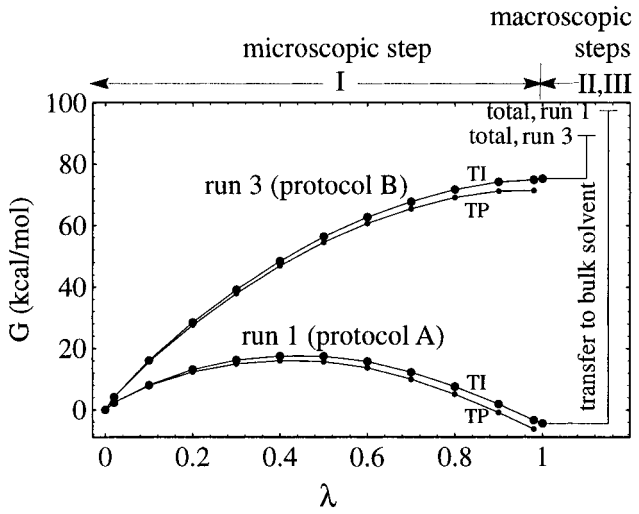

Figure 6. Free energy from the same two runs (1 and 3) shown in Figures 4 and 5, which used scaling protocols A and B, respectively. Thermodynamic integration (TI) and perturbation (TP) results are shown. These microscopic free energies represent step I of the cycle (see Figure 1). The continuum corrections, from steps II and III, are shown on the right as solid vertical lines for each protocol ("transfer to bulk solvent").

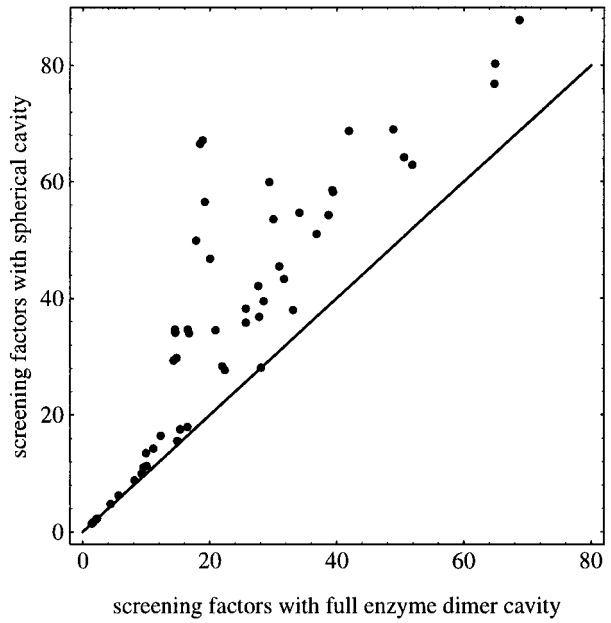

Figure 7. Potential-based screening factors $1/\alpha_i = V_v(i\rightarrow 0)/V_w(i\rightarrow 0)$ using either the spherical dielectric cavity or the full enzyme dimer as the microscopic cavity.

9.5 kcal/mol self-energy, and the six unscaled residues not involved in salt bridges contribute the same 4.5 kcal/mol.

In protocol C, a different set of seven additional residues have been scaled, allowing us to test the sensitivity of the results to the exact way the system is neutralized. The Asp ligand and the six unscaled residues contribute as above. Four additional residues contribute 10.5 kcal/mol (Arg174 and Arg141 are no longer in salt bridges, since their partners are now scaled; Glu171 and Asp282 are not scaled in this protocol).

The continuum contributions in protocols A, B, and C, compensate almost exactly the large differences in the microscopic free energy results. Thus the overall free energies, reported in Table 4, are in reasonable agreement, varying between 91.0 and 98.2 kcal/mol for five different runs. The runs shown in Figure 6 agree within 6.6 kcal/mol, while the continuum corrections in the two protocols (A and B) differ by 88 kcal/mol (see Table 3). In other words, the *a priori* scaling of different selected charges and their *a posteriori* screening by bulk solvent give very similar results. This suggests that the potential-based charge scaling is a robust way to correct for bulk solvent screening, which is also theoretically correct in the context of continuum electrostatics.

4.5. Continuum Calculations with Spherical Reaction Field Treatment of Solvent. The continuum contributions in steps II and III were also calculated, for comparison, using a cavity that contains only the microscopic model, as opposed to the complete enzyme dimer. The microscopic model is roughly spherical, so that this calculation approximates a spherical reaction field treatment of the solvent.[24−26]

The use of the microscopic model modifies the screening of the protein charges. Low-dielectric protein outside the microscopic region is replaced by high dielectric solvent, and so the potential-based scaling factors of the full enzyme dimer model are replaced by systematically larger scaling factors. The two sets are compared in Figure 7. The agreement is quite good in the inner half of the microscopic region (less than 10 Å from the ligand), but is quite poor in the outer half. However, this disagreement has little effect on the free energy difference, ΔG(Asp→Asn): going from the full enzyme dimer to the spherical dielectric cavity only changes the free energy difference by +1.1 kcal/mol. This increase is due to the solvation by the additional bulk solvent of the Asp charge, and of the predominantly negative protein charges located more than 10 Å from the ligand (Figure 3). The predominantly positive charges located in the innermost 10 Å are only slightly screened by the additional solvent. However the overall effect is small, because the residues in the outermost 10 Å are already strongly screened by bulk solvent in the full dimer model; and because the self-energy of the Asp (9.5 kcal/mol) is already close to the value for a spherical cavity and cannot increase significantly.

The spherical approximation is expected to modify the dynamics of the system somewhat, since the scaling factors obtained by its use are significantly different in parts of the simulation system. Consequently, the structures and the free energy obtained in step I will also be affected. We have

compared the solvent reaction field on the ligand and protein atoms, using the spherical dielectric cavity, to the exact values, using the full enzyme dimer. For atoms in the inner half of the microscopic region (less than 10 Å from the ligand C_γ), the electric field with the spherical approximation deviates by just 5% on average from the exact field. For the outer half, the average deviation is 40%. Thus while the agreement is very good in the ligand region, it is poor in the outer part of the microscopic region. To evaluate the exact effect on the dynamics further, additional simulations are needed, and systems with other geometries must be investigated.

5. Concluding Discussion

Free energy simulations for processes that add or delete net charges in proteins are difficult, and a rather limited number of examples exist in the literature.[22,32,33,38,65,83−85] The essential problem is the long range of the electrostatic interaction, and the need to limit the size of the system used for a detailed free energy simulation. Here we propose a method to perform such calculations, which is related to the methods proposed by Resat and McCammon,[37] by Prévost,[38] and by Fischer *et al*.[30] for a neutral ligand. The calculation is divided into a three-step process (Figure 1): a mutation step, a charge unscaling step, and a transfer step into bulk water. Solvent and protein outside a microscopic region are treated by continuum electrostatics. For a microscopic region with a 20 Å radius, such as the one considered in the application above, this approximation should be very good. With it, all long-range electrostatic interactions are included in the calculation. In addition, it is possible to include all electrostatic interactions within the microscopic region as well, for example through a multipole treatment, as in the present application. The dielectric inhomogeneity of the system outside the microscopic region is taken explicitly into account through the numerical solution of the Poisson−Boltzman equation, which includes the exact shape of the entire protein. The detailed dielectric relaxation of charged groups near the mutation site in response to the mutation is explicitly modeled at a microscopic level by the molecular dynamics simulations.

The method is very simple to implement. It requires only the precalculation of the electrostatic potential at the mutation site with a continuum model for the simulation system surrounded by vacuum and bulk water, respectively. The method has been shown to be robust with respect to the exact details of the charge scaling protocol. This confirms that the continuum treatment of the protein and solvent is a good approximation. Indeed, despite the very large differences between the continuum contributions associated with three different charge scaling protocols (on the order of 90 kcal/mol, compared to the overall free energy change of 95 kcal/mol), the results of five free energy runs were in good agreement. Thus, the macroscopic contributions of one protocol were compensated by microscopic contributions in the other. The method should be applicable to any process that modifies a localized net charge or group of charges in a protein. This includes calculations of pK_a's of titratable sites, redox potentials, or binding free energy differences between charged and neutral ligands.

Several alternative approaches exist. A fully macroscopic treatment could be used, where protein and solvent are viewed as two (possibly inhomogeneous) macroscopic dielectric continua, and the Poisson−Boltzmann equation solved. Results with this approach for the present application will be described elsewhere. They demonstrate that the detailed structural relaxation and the large free energy contributions of charged groups near the ligand are not accurately modelled by a pure continuum approach. The Langevin dipole method[16,22] is expected to exhibit similar limitations.

Another possibility is to use a spherical reaction field approximation for the long-range electrostatics, which treats the region outside the microscopic region as homogeneous high-dielectric solvent. With the spherical approximation, the reaction field can be calculated "on the fly" during the molecular dynamics simulation.[24−26] However, the inhomogeneity of the external region is neglected. While this has only a small effect on the bulk solvent contribution to binding (in the present application), it presumably does modify the dynamics and the free energy for the microscopic mutation step. The importance of this effect remains to be tested.

To compare quantitatively the binding affinities of the substrate aspartic acid and the substrate analogue asparagine for aspartyl-tRNA synthetase, we have also carried out the Asp→Asn mutation in solution. With the calculations reported here, we are then able to make a detailed analysis of the interactions involved in specific aspartyl-tRNA synthetase−substrate recognition and make predictions as to the effect of mutations on binding. These results will be reported elsewhere.[41]

Acknowledgment. We thank Michael Schaefer for performing some of the UHBD calculations. M.K. acknowledges partial support by the National Institutes of Health, the CNRS, and the Ministère de l'Education. Some of the calculations were performed at the IDRIS supercomputer center of the CNRS (grants to M.K. and T.S.).

References and Notes

(1) Tanford, C.; Kirkwood, J. *J. Am. Chem. Soc.* **1957**, *79*, 5333.
(2) Warwicker, J.; Watson, H. *J. Mol. Biol.* **1982**, *157*, 671.
(3) Rogers, N. *Prog. Biophys. Mol. Biol.* **1986**, *48*, 37−66.
(4) Sharp, K.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *19*, 301−332.
(5) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(6) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219−10225.
(7) Bashford, D.; Case, D.; Dalvit, C.; Tennant, L.; Wright, P. *Biochemistry* **1993**, *32*, 8045−8056.
(8) Antosiewicz, J.; McCammon, J.; Gilson, M. *J. Mol. Biol.* **1994**, *238*, 415−436.
(9) Simonson, T.; Perahia, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082−1086.
(10) Simonson, T.; Perahia, D. *J. Am. Chem. Soc.* **1995**, *117*, 7987−8000.
(11) Simonson, T.; Brooks, C. L., III. *J. Am. Chem. Soc.* **1996**, *118*, 8452−8458.
(12) Northrup, S.; Wensel, T.; Meares, C.; Wendolowski, J.; Matthew, J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9503−9507.
(13) Pack, G.; Garrett, G.; Wong, L.; Lamm, G. *Protein Eng.* **1993**, *2*, 177−183.
(14) Chan, S.; Lim, C. *J. Phys. Chem.* **1994**, *98*, 692−695.
(15) You, T.; Bashford, D. *Biophys. J.* **1996**, *69*, 1721−1733.
(16) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(17) Simonson, T.; Perahia, D.; Bricogne, G. *J. Mol. Biol.* **1991**, *218*, 859−886.
(18) Warshel, A.; Russell, S. *Q. Rev. Biophys.* **1984**, *17*, 283−342.
(19) Mezei, M.; Beveridge, D. Free energy simulations. In *Computer Simulation of Chemical and Biomolecular Systems*; Beveridge, D., Jorgensen, W., Eds.; New York Academy of Sciences: New York, 1986; Vol. 482, pp 1−23.
(20) Brooks, C., III; Karplus, M.; Pettitt, M. *Adv. Chem. Phys.* **1987**, *71*, 1−259.
(21) McCammon, J.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.
(22) Warshel, A. *Computer Modelling of Chemical Reactions in Enzymes and Solutions*; John Wiley: New York, 1991.
(23) York, D.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. *J. Am. Chem. Soc.* **1995**, *117*, 5001−5002.
(24) Alper, H.; Levy, R. *J. Chem. Phys.* **1993**, *99*, 9847−9852.
(25) Beglov, D.; Roux, B. *J. Phys. Chem.* **1994**, *96*, 6428−6431.
(26) Wang, L.; Hermans, J. *J. Phys. Chem.* **1995**, *99*, 12001−12007.
(27) Brooks, C., III; Brünger, A. T.; Karplus, M. *Biopolymers* **1985**, *24*, 843−865.

(28) Brooks, C., III; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159−181.
(29) Stote, R.; States, D.; Karplus, M. *J. Chim. Phys.* **1991**, *88*, 2419−2433.
(30) Fischer, S.; Michnick, S.; Karplus, M. *Biochemistry* **1991**, *32*, 13830−13837.
(31) Gunsteren, W. v.; Berendsen, H. *Groningen Molecular Simulation Library Manual*; *Biomos*: Groningen, The Netherlands, 1987.
(32) Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368.
(33) Hwang, J.-K.; Warshel, A. *Biochemistry* **1987**, *26*, 2669.
(34) Gilson, M.; Davis, M.; Luty, B.; McCammon, J. *J. Phys. Chem.* **1993**, *97*, 3591−3600.
(35) Juffer, A. On the Modelling of Solvent Mean Force Potentials. From Liquid Argon to Macromolecules. Ph.D. Thesis, University of Groningen, The Netherlands, 1993.
(36) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(37) Resat, H.; McCammon, J. *J. Chem. Phys.* **1996**, *104*, 7645−7651.
(38) Prévost, M. *J. Mol. Biol.* **1996**, *260*, 99−110.
(39) Straatsma, T. P. Free Energy Evaluation by Molecular Dynamics Simulation. Ph.D. Thesis, University of Groningen, The Netherlands, 1987.
(40) Tembe, B.; McCammon, J. *Comput. Chem.* **1984**, *8*, 281−283.
(41) Archontis, G.; Simonson, T.; Moras, D.; Karplus, M. *J. Mol. Biol.* (in press).
(42) Allen, M.; Tildesley, D. *Computer Simulations of Liquids*; Clarendon Press: Oxford, 1991.
(43) Kirkwood, J. *J. Chem. Phys.* **1935**, *3*, 300−313.
(44) Fowler, R. H.; Guggenheim, E. A. *Statistical Thermodynamics*; Cambridge University Press: Cambridge, 1939.
(45) Press, W.; Flannery, B.; Teukolsky, S.; Vetterling, W. *Numerical Recipes*; Cambridge University Press: Cambridge, 1986.
(46) Hummer, G.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 2004−2010.
(47) Friedman, H. *Mol. Phys.* **1975**, *29*, 1533−1543.
(48) Landau, L.; Lifschitz, E. *Electrodynamics of Continuous Media*; Pergamon Press: New York, 1980.
(49) Eiler, S. Ph.D. Thesis, University of Strasbourg, France, 1997.
(50) Poterszman, A.; Delarue, M.; Thierry, J.; Moras, D. *J. Mol. Biol.* **1994**, *244*, 158−167.
(51) Brünger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148−156.
(52) Simonson, T. *Chem. Phys. Lett.* **1996**, *250*, 450−454.
(53) Gilson, M.; Sharp, K.; Honig, B. *J. Comput. Chem.* **1988**, *9*, 327−335.
(54) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−445.
(55) Davis, M.; McCammon, J. *J. Comput. Chem.* **1991**, *7*, 909.
(56) Davis, M.; Madura, J.; Luty, B.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187.
(57) Mackerell, A.; et al. *J. Phys. Chem.* (submitted).
(58) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(59) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(60) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and molecular dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(61) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327−341.
(62) Brooks, C., III; Karplus, M. *J. Chem. Phys.* **1984**, *79*, 6312−6325.
(63) Tidor, B.; Karplus, M. *Biochemistry* **1991**, *30*, 3217−3228.
(64) Tidor, B. Unpublished work.
(65) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. *Science* **1989**, *244*, 1069−1072.
(66) Boresch, S.; Karplus, M. In preparation.
(67) Simonson, T. *Mol. Phys.* **1993**, *80*, 441−447.
(68) Resat, H.; Mezei, M. *J. Chem. Phys.* **1994**, *222*, 6126−6140.
(69) Beutler, T.; Mark, A.; Schaik, R. v.; Gerber, P.; Gunsteren, W. v. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(70) Demailly, J. *Analyse Numérique et Équations Différentielles*; Presses Universitaires de Grenoble: Grenoble, France.
(71) Hermans, J. *J. Phys. Chem.* **1991**, *95*, 9029−9032.
(72) Wood, R. H. *J. Phys. Chem.* **1991**, *95*, 4838−4842.
(73) Picinbono, B. *Phys. Rev. A* **1977**, *16*, 2174−2177.
(74) Valleau, J.; Torrie, G. A guide to Monte Carlo for statistical mechanics. In *Modern Theoretical Chemistry*; Berne, B., Ed.; Plenum Press: New York, 1977; Vol. 5.
(75) Allen, M.; Tildesley, D. *Computer Simulations of Liquids*; Clarendon Press: Oxford, 1991.
(76) Schiferl, S.; Wallace, D. *J. Chem. Phys.* **1985**, *83*, 5203−5209.
(77) McQuarrie, D. *Statistical Mechanics*; Harper and Row: New York, 1975.
(78) Tobias, D. J.; Brooks, C. L., III; Fleischman, S. H. *Chem. Phys. Lett.* **1989**, *156*, 256−260.
(79) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1989**, *90*, 3300−3304.
(80) Hermans, J.; Yun, R. H.; Anderson, A. G. *J. Comput. Chem.* **1992**, *13*, 429−442.
(81) Wade, R.; McCammon, J. *J. Mol. Biol.* **1992**, *225*, 679−712.
(82) Hodel, A.; Simonson, T.; Fox, R. O.; Brünger, A. T. *J. Phys. Chem.* **1993**, *97*, 3409−3417.
(83) Kuczera, K.; Gao, J.; Tidor, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8481−8485.
(84) Prod'hom, B.; Karplus, M. *Protein Eng.* **1993**, *6*, 585−592.
(85) Aqvist, J. *J. Comput. Chem.* **1996**, *17*, 1587−1597.