

# Electrostatic free energy calculations using the generalized solvent boundary potential method

Nilesh K. Banavali, Wonpil Im, and Benoît Roux<sup>a)</sup>

*Department of Biochemistry and Structural Biology, Weill Medical College of Cornell University,  
New York, New York 10021*

(Received 28 May 2002; accepted 23 July 2002)

Free energy perturbation (FEP) calculations using all-atom molecular dynamics simulations with a large number of explicit solvent molecules are a powerful approach to study ligand–macromolecule association processes at the atomic level. One strategy to carry out FEP calculations efficiently and reduce computational time is to consider the explicit dynamics of only a small number of atoms in a localized region around the ligand. Such an approximation is motivated by the observation that the factors governing binding specificity are dominated by interactions in the vicinity of the ligand. However, a straightforward truncation of the system may yield inaccurate results as the influence exerted by the remote regions of the macromolecule and the surrounding solvent through long-range electrostatic effects may be significant. To obtain meaningful results, it is important to incorporate the influence of the remote regions of the ligand–macromolecule complex implicitly using some effective potential. The generalized solvent boundary potential (GSBP) that was developed recently [W. Im, S. Bernèche, and B. Roux, *J. Chem. Phys.* **114**, 2924 (2001)] is an efficient computational method to represent the long-range electrostatic interactions arising from remote (outer) regions in simulations of a localized (inner) region with a small number of explicit atoms. In the present work, FEP calculations combined with GSBP are used to illustrate the importance of these long-range electrostatic factors in estimation of the charging free energy of an aspartate ligand bound to the aspartyl-tRNA synthetase. Calculations with explicit spherical simulation inner regions of different radii are used to test the accuracy of the GSBP method and also illustrate the importance of explicit protein and solvent dynamics in the free energy estimation. The influence of the represented outer region is tested using separate simulations in which the reaction field and/or the protein static field are excluded. Both components are shown to be essential to obtain quantitatively meaningful results. The ability of implicitly treating the influence of protein fluctuations in the outer region using a protein dielectric constant is examined. It is shown that accurate charging free energy calculations can be performed for this system with a spherical region of 15 to 20 Å radius, which roughly corresponds to 1500–3500 moving atoms. The results indicate that GSBP in combination with FEP calculations is a precise and efficient approach to include long-range electrostatic effects in the study of ligand binding to large macromolecules. © 2002 American Institute of Physics.  
[DOI: 10.1063/1.1507108]

## I. INTRODUCTION

Molecular recognition phenomena that require the association of ligands to macromolecules with high affinity and specificity play a key role in biology.<sup>1–3</sup> An important quantitative link between experimental measurements and theoretical estimations based on atomic models for binding processes of ligands to macromolecules is provided by the binding free energy.<sup>4,5</sup> Predicting the binding free energy of ligands to a macromolecule can also have great pragmatic value in identifying novel molecules that can bind to target receptors and act as therapeutic drugs.<sup>6</sup> Although the fundamental microscopic interactions giving rise to molecular recognition are relatively well-understood, designing computational schemes to accurately calculate binding free energies remains very challenging. Computational approaches used

for screening large molecular databases of compounds to identify potential lead drug molecules must rely on very simplified approximations in order to achieve the needed computational efficiency.<sup>7</sup> However, since the binding free energies are typically on the order of a few kcal/mol, the calculated free energies must be very accurate to have any predictive value.

Molecular dynamics (MD) simulations with free energy perturbation (FEP) methods are arguably the most powerful and promising approach to estimate the binding free energies of ligands to macromolecules using atomic models.<sup>8–13</sup> FEP calculations, however, are intrinsically limited by several difficulties which range from systematic bias due to force field inaccuracies and insufficient configurational sampling.<sup>12,14</sup> In addition, the treatment of long-range electrostatic interactions is of particular importance in the case of FEP calculations involving charged species.<sup>15</sup> Despite the recent developments in MD methodologies,<sup>16</sup> carrying out FEP

<sup>a)</sup>Telephone: 212-746-6018; fax: 212-746-4843; electronic mail: Benoit.Roux@med.cornell.edu

calculations of large macromolecular assemblies surrounded by explicit solvent molecules often remains computationally prohibitive. For this reason, it is necessary to seek ways to reduce the computational cost of FEP calculations. Binding specificity is often dominated by local interactions in the vicinity of the ligand while remote regions are expected to contribute in an average manner. Accordingly, a hybrid approach in which only a small region of the system around the ligand is explicitly simulated while the influence of the remote parts is represented implicitly using an effective potential can be an attractive strategy to study the association of ligands to macromolecular complexes.

Methods to incorporate the influence of the surrounding on a small subset of atoms embedded in a large macromolecule go back to the early simulations of active sites of proteins in an explicit solvent environment.<sup>17,18</sup> In these early studies, the system was typically truncated beyond a certain distance away from the region of interest to reduce the computational cost.<sup>19,20</sup> Such a truncation, however, neglects or incorrectly estimates long-range electrostatic forces exerted by the remote parts of the macromolecule and the surrounding solvent which may contribute significantly.<sup>15,21</sup> Several efforts have been made to address this issue, and a number of computational schemes have been proposed to permit accurate FEP calculations.<sup>15,21–25</sup> Recently, a generalized solvent boundary potential (GSBP) method has been developed for accurately simulating a small subset of atoms embedded in a large solvated macromolecular system of arbitrary geometry.<sup>26</sup> GSBP is a generalization of the spherical solvent boundary potential (SSBP) which was previously developed to simulate the isotropic bulk liquid surrounding a solute.<sup>27</sup> In the GSBP method, all atoms in the inner region belonging to either ligand, macromolecule, or solvent can undergo explicit dynamics whereas the macromolecular and solvent atoms outside this inner region are included implicitly. The influence of the surrounding outer region of the macromolecular system is represented in terms of a solvent-shielded static field and a solvent-induced reaction field. The reaction field due to changes in charge distribution in the dynamic inner region is expressed in terms of a basis set expansion of the inner simulation region charge density, the basis set coefficients corresponding to generalized electrostatic multipoles. The solvent-shielded static field from outer macromolecular atoms and the reaction field matrix representing the couplings between the generalized multipoles are both invariant with respect to the configuration of the explicit atoms in the inner simulation region. They are calculated only once for macromolecules of arbitrary geometry using the finite-difference Poisson–Boltzmann (PB) equation, leading to an accurate and computationally efficient hybrid MD/continuum method for simulating a small region of a large biological macromolecular system.

Our goal in the present work is to assess the ability of GSBP to incorporate the long-range electrostatic interactions accurately in free energy calculations involving a charged ligand bound to a macromolecule. The performance of GSBP in accurately determining the charging free energy of a zwitterionic aspartic acid residue in a specific conformation in the active site of the aspartyl-tRNA synthetase dimer is

tested with FEP calculations. Aspartyl tRNA synthetase is a large macromolecule that catalyzes conjugation of an aspartate to its corresponding tRNA to generate the activated aspartyl-tRNA that is then utilized in protein synthesis.<sup>28–32</sup> Ensuring amino acid specificity while catalyzing this conjugation is of the utmost importance in preventing potentially harmful mutations in the synthesized protein.<sup>28</sup> The ability of the aspartyl-tRNA synthetase to distinguish between a charged aspartate and a neutral asparagine ligand is therefore a good example of the ability of the protein to distinguish between ligands based on their electrostatic properties. Previous theoretical studies have carefully characterized the free energy difference between binding of the aspartate ligand and the asparagine ligand using different approaches.<sup>15,22–25,33</sup> For this reason, this system is a good testing ground for GSBP. The present calculations are used to address the limited and well-defined problem of estimating the electrostatic component of the binding free energy of the aspartate to the aspartyl-tRNA synthetase in a fixed orientation. Even though the difference in binding free energy between aspartate and asparagine is a more biologically relevant quantity, some difficulties arise due to the lack of experimental data on the binding conformation of the asparagine and the fact that it may differ significantly from the binding conformation of the aspartate.<sup>15,22–25,33</sup> In order to assess the ability of GSBP to properly treat the electrostatic contribution in FEP calculations, it is desirable to avoid these additional complexities. Therefore the charging free energy of the aspartate ligand both bound to the protein and in solvent is calculated with GSBP using inner regions of different sizes and the contribution from various components is analyzed in detail. On the basis of these calculations, it is concluded that GSBP is a computationally efficient method for accurately estimating electrostatic free energies in large macromolecular systems.

## II. COMPUTATIONAL DETAILS

The GSBP method<sup>26</sup> was implemented in the PBEQ module<sup>34–36</sup> of the biomolecular simulation program CHARMM.<sup>37</sup> The total effective potential energy  $\mathcal{W}_{\text{GSBP}}$  for the simulation can be written as<sup>26</sup>

$$\mathcal{W}_{\text{GSBP}} = U^{(\text{ii})} + U_{\text{int}}^{(\text{io})} + U_{\text{LJ}}^{(\text{io})} + \Delta\mathcal{W}_{\text{np}} + \Delta\mathcal{W}_{\text{sf}} + \Delta\mathcal{W}_{\text{rf}}, \quad (1)$$

where  $U^{(\text{ii})}$  is the total inner–inner potential energy, and  $U_{\text{int}}^{(\text{io})}$  and  $U_{\text{LJ}}^{(\text{io})}$  are, respectively, the internal (bonds, angles, and torsions) and the Lennard-Jones contributions to the inner–outer potential energy, and  $\Delta\mathcal{W}_{\text{np}}$  is a nonpolar confining potential.<sup>19,27</sup> The last two terms in Eq. (1) represent the long-range electrostatic contribution from the outer region onto the inner region. These two terms correspond to the solvent-shielded static field,

$$\mathcal{W}_{\text{sf}} = \sum_{\alpha \in \text{inner}} q_{\alpha} \phi_s^{(\text{o})}(\mathbf{r}_{\alpha}) \quad (2)$$

and the reaction field

$$\mathcal{W}_{\text{rf}} = \frac{1}{2} \sum_{mn} Q_m M_{mn} Q_n \quad (3)$$

[in the initial theoretical developments leading to Eq. (3)<sup>26</sup> the reaction field term was called  $\Delta\mathcal{W}_{\text{elec}}^{(\text{ii})}$  in reference to inner-inner electrostatic interactions].

In GSBP, the electrostatic potential  $\phi_s^{(o)}(\mathbf{r})$  and the generalized reaction field matrix  $\mathbf{M}$  are first calculated, and then stored for simulations. To calculate  $\phi_s^{(o)}(\mathbf{r})$ , the PB equation for the system,

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi_s^{(o)}(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r}) \phi_s^{(o)}(\mathbf{r}) = -4\pi\rho^{(o)}(\mathbf{r}), \quad (4)$$

is numerically solved using a finite-difference method<sup>38,39</sup> with the charge density  $\rho^{(o)}(\mathbf{r})$  from the outer region (all atomic charges in the inner region are set to zero). The result of the calculation, i.e., the electrostatic potential on each point of the discrete grid, is then stored for the simulations (the subscript “s” indicates that the potential  $\phi$  is calculated with the full solvent dielectric environment). To calculate the generalized reaction field matrix  $\mathbf{M}$ , the PB equation is solved using the basis function as a continuous charge distribution in the extended dielectric cavity with the full dielectric environment,

$$\begin{aligned} \nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi_s(\mathbf{r}; b_m(\mathbf{r}))] - \bar{\kappa}^2(\mathbf{r}) \phi_s(\mathbf{r}; b_m(\mathbf{r})) \\ = -4\pi b_m(\mathbf{r}) \end{aligned} \quad (5)$$

(all explicit atomic charges in the system are set to zero in this calculation). In the present application, spherical harmonic functions were used for a spherical cavity corresponding to the radius of the inner region extended by  $\Delta R_{\text{diel}}$  of 3 Å. Spherical inner region of radius 10, 15, and 20 Å centered on the ligand were considered.

The  $nm$ th matrix element corresponds to the interaction between,  $\phi_{\text{rf}}(\mathbf{r}; b_m(\mathbf{r}))$ , the reaction field due to the charge density supported by the basis function  $b_m(\mathbf{r})$ , and the charge density of  $b_n(\mathbf{r})$ ,

$$M_{nm} = \int d\mathbf{r} b_n(\mathbf{r}) \phi_{\text{rf}}(\mathbf{r}; b_m(\mathbf{r})). \quad (6)$$

The reaction field due to the basis function  $b_m(\mathbf{r})$  is calculated by subtracting the electrostatic potential from the reference vacuum system,<sup>26</sup>  $\phi_{\text{rf}}(\mathbf{r}; b_m(\mathbf{r})) \equiv [\phi_s(\mathbf{r}; b_m(\mathbf{r})) - \phi_v(\mathbf{r}; b_m(\mathbf{r}))]$ , with

$$\nabla \cdot [\nabla \phi_v(\mathbf{r}; b_m(\mathbf{r}))] = -4\pi b_m(\mathbf{r}) \quad (7)$$

(the subscript “v” indicates that the potential  $\phi$  is calculated with the vacuum dielectric environment).

The dielectric constant of the solvent region was set to 80 (note that the dielectric constant of the solvent region is set to 1 for the vacuum calculation). No salt was included and the ionic screening constant  $\bar{\kappa}^2(\mathbf{r})$  was set to 0 everywhere. Different values of the dielectric constant of the protein in the outer region ( $\epsilon_{\text{prot}}$ ), varying from 1 to 4, were considered. The purpose of varying the outer region dielectric constant is to incorporate the influence of the dynamical fluctuations of the protein region on the charging free energy. The protein-solvent boundary was set using the optimized atomic born radii,<sup>34</sup> including the reentrant of the molecular surface determined with a probe radius of 1.4 Å. To reduce the number of grid-charges in setting up the boundary conditions at the edge of the grid, the PB equation is first solved

for a coarse grid with large spatial extent and the result is refined subsequently by using a focused calculation.<sup>40</sup>

The initial atomic model of the aspartate ligand bound to the aspartyl-tRNA synthetase was graciously provided by Tom Simonson.<sup>15,22</sup> To set up the FEP calculation, the region of the complex surrounding the aspartate ligand was solvated using a preequilibrated solvent sphere of radius 25 Å. The sphere of solvent molecules was centered on the aspartate ligand and all solvent molecules within 2.6 Å of any nonhydrogen protein atom were removed. Successive overlays<sup>41</sup> of the solvent sphere were performed with 10 ps equilibration each with the solute atoms held fixed followed by removal of newly added solvent molecules within 2.6 Å of previously existing nonhydrogen solute and solvent atoms. These iterative overlays were performed until the number of solvent molecules added in successive overlays converged to a number less than 5. Water molecules within 3 Å of any solute atom but not within 3.5 Å of other solvent atoms were removed to avoid introducing spurious water molecules into the protein interior. After deletion of all water molecules beyond a 20 Å radius the final structure obtained was equilibrated for 550 ps with the solute and solvent atoms in the inner region of radius 20 Å being unconstrained. The final equilibrated structure was then used to initiate the FEP calculations. The smaller simulation systems of radii 10 Å and 15 Å were generated directly from the larger sphere structure by merely deleting excess outer explicit solvent molecules. The total number of atoms in these systems is 468, 1592, and 3717, and the total number of water molecules is 36, 105, and 1071, for the 10, 15, and 20 Å inner region radius systems, respectively. For comparison, the FEP calculations were repeated for the same configuration of the aspartate ligand in isotropic bulk water spherical systems with GSBP (there is no static field in this case). Those bulk water systems have a radius of 10, 15, and 20 Å, and contain 136, 473, and 1142 water molecules, respectively.

All MD simulations were carried out using Langevin dynamics at 300 K with a friction constant corresponding to a relaxation time of 5 ps<sup>-1</sup> applied to the nonhydrogen atoms. The water geometry was kept fixed using SHAKE<sup>42</sup> and the integration timestep used was 0.002 ps. No cutoff was applied for the inner region electrostatic interactions but all electrostatic interactions beyond 12 Å were treated on the basis of dipolar and quadrupolar expansions using the extended electrostatic method.<sup>43</sup> This reduced the computational time of the calculation by about a factor of 2 relative to an infinite cutoff scheme in the present simulations. During the simulations, the generalized multipole moments  $Q_n$  in Eq. (1) were calculated using the spherical harmonic basis functions for every instantaneous configuration of the atoms in the inner region. All the atoms (macromolecule, ligand, and solvent) lying in the inner region were allowed to move except the macromolecule atoms in the intermediate space at the boundary between the inner and outer regions. Atoms near the boundary of the spherical system were fixed according to a group-based criteria. These fixed macromolecular atoms were assigned to the inner region and were located in the extended spherical dielectric cavity whose radius is greater than the inner region radius by  $\Delta R_{\text{diel}}$  (3.0 Å). This



TABLE I. FEP calculations with fixed protein (results in kcal/mol). A dielectric constant of 1.0 was used for the outer protein region.

Calculation	Radius of the inner region		
	10 Å	15 Å	20 Å
Including all interactions	172.1 (0.3)	174.4 (0.7)	168.4 (0.7)
Simple truncated system	226.6	122.3	146.8
Excluding static field	202.2	173.4	177.3
Excluding reaction field	202.8	123.7	140.7
Protein charges turned off	64.5	68.4	71.6

procedure facilitates the correct handling of the 1–2 and 1–3 nonbonded exclusion between covalently bonded neighbors of the inner and outer regions.<sup>26</sup> The fixed protein atom nearest to the aspartate ligand was 7.3, 12.4, and 16.1 Å away from the inner region sphere center for the 10, 15, and 20 Å radius spheres, respectively.

The FEP calculations were carried out using the PERT module in CHARMM which was modified to enable the use of GSBP and Extended Electrostatics<sup>43</sup> to treat the nonbonded interactions between inner region atoms. The value of  $\lambda$  was varied from 0 to 1 in 10 steps with 100 steps of ABNR minimization followed by 5 ps equilibration and 20 ps sampling for each window in the forward and reverse directions, yielding a total sampling time of 440 ps per FEP simulation. Five separate simulations were carried out for each FEP calculations where standard deviations are reported. The free energy change ( $\Delta G$ ) was computed using the WHAM algorithm<sup>44,45</sup> to iteratively determine the free energy difference between the initial and final state of the perturbation. For each of the five simulations,  $\Delta G$  was computed separately by the WHAM algorithm and the average and standard deviation of these values were determined.

### III. RESULTS AND DISCUSSION

To examine the numerical accuracy of the GSBP static and reaction field, we first examined the charging free energy with the protein and the aspartic acid ligand held in a fixed configuration. Table I shows the FEP calculations with different electrostatic components either included or excluded. It is observed that the charging free energy calculated with all GSBP terms (static and reaction field) yields results that are nearly independent of the size of the inner region. The free energy change going from charged to uncharged state of the aspartate ligand is 172 kcal/mol on average with variations on the order of about 5 kcal/mol, which is less than 3 percent of the total change. In sharp contrast, ignoring both the outer protein static field and outer region reaction field by using a truncated sphere consisting of just the inner region atoms yields grossly incorrect values for all sphere sizes. This illustrates the significant impact that long-range electrostatic interactions have on the free energy of a charged ligand. FEP calculations performed while excluding the static field or the reaction field contributions yield also incorrect results, which demonstrates that each of these components is essential for an accurate treatment of the boundary potential. The exclusion of the outer protein static field causes variations in the calculated free energy for the system

TABLE II. FEP calculations in bulk water (results in kcal/mol). Standard deviations estimated from 5 independent runs are shown in parentheses.

Radius of the inner region		
10 Å	15 Å	20 Å
111.1 (0.3)	112.4 (0.3)	112.8 (0.4)

of different sizes. The lack of a clear trend in the change is not surprising since the protein regions which are excluded are very different for the different sphere sizes. The exclusion of the reaction field response from the outer region due to the inner region charges also affects all calculated free energy values to a large extent (more than 15 kcal/mol). It is particularly noteworthy that increasing the size of the explicit inner region cannot improve the results to a significant degree if the reaction field is not included. In earlier stochastic boundary simulations,<sup>20,46</sup> these contributions as well as the shielding of the outer region protein static field were typically ignored.

In the present calculations, the solvent-shielded static field component corresponds to the numerical representation of the electrostatic potential  $\phi_s^{(o)}(\mathbf{r})$  arising from a fixed protein stored on a discrete grid. Previous tests have shown that such a numerical procedure is very accurate when a grid spacing of 0.5 Å is used.<sup>26</sup> On the other hand, the reaction field is represented in terms of a generalized multipolar expansion which relies on a basis set expansion. The coupling coefficients between the generalized multipoles are extracted from a series of finite-difference PB calculations using Eqs. (6) and (7). To examine the accuracy of this approximate procedure, FEP calculations were performed with all protein charges turned off. These calculations yield the charging free energy of the aspartate ligand interacting only with the explicit water molecules in the inner region and the high dielectric continuum representing the water molecules in the outer region (i.e., the reaction field contribution). The results are 64.5, 68.4, and 71.6 kcal/mol for the inner region of 10, 15, and 20 Å radius, respectively, indicating that there are moderate variations with system size in the accuracy of the reaction field component represented in terms of a generalized basis set expansion. In comparison, the variations in the charging free energy given in Table II are much smaller in the case of an isotropic bulk water environment (with no protein). This is probably due to the fact that the demands on the generalized basis set expansion are much greater when the reaction field is representing a strongly anisotropic environment such as a protein binding site rather than an isotropic bulk solution. Furthermore, it should be stressed that the aspartyl-tRNA synthetase simulation system of 10 Å radius contains only 36 explicit water molecules, compared to 136 water molecules in the corresponding 10 Å radius bulk water system. Interestingly, the variations in the total charging free energy given in Table I are nevertheless quite small, suggesting that there is some compensation and cancellation of errors when all the terms in GSBP are included. The convergence of the final result given such limited explicit representation of water in the smallest system is quite remarkable.

TABLE III. FEP calculations with flexible protein (results in kcal/mol). Standard deviation estimated from 5 independent runs are shown in parentheses.

Calculation $\epsilon_{\text{prot}}$	Radius of the inner region		
	10 Å	15 Å	20 Å
Including all interactions ( $\epsilon_{\text{prot}}=1$ )	127.6 (0.6)	129.2 (0.4)	107.4 (2.1)
Including all interactions ( $\epsilon_{\text{prot}}=2$ )	146.1 (0.5)	116.1 (0.9)	110.5 (2.0)
Including all interactions ( $\epsilon_{\text{prot}}=3$ )	153.6 (0.7)	110.6 (1.4)	107.9 (1.5)
Including all interactions ( $\epsilon_{\text{prot}}=4$ )	155.8 (1.0)	108.8 (1.2)	108.8 (2.0)
Excluding static field ( $\epsilon_{\text{prot}}=4$ )	168.9	110.1	114.8
Excluding reaction field ( $\epsilon_{\text{prot}}=4$ )	175.6	77.7	84.3
Truncated system	184.7	76.3	88.2

In all the calculations mentioned above, the outer protein region was assigned a dielectric constant of 1. This choice was made primarily to allow a direct comparison of the results from FEP calculations for simulated systems of different size. In the absence of induced electronic polarization, fixed protein atoms do not generate any reaction field, and the correspondence to a dielectric constant of 1 is rigorously correct. In reality, of course, proteins are not rigid and it is important to consider the influence of dynamical flexibility. These effects are likely to be complex and somewhat system dependent. Generally, it may be expected that protein fluctuations in the outer region will affect both the static and reaction fields. For example, the average field arising from a dynamically fluctuating protein in the outer region might be smaller than the static field arising from fixed protein atoms. Correspondingly, the average orientation of the polar moieties in the outer region might be affected in response to the field from the charged ligand, which in turn give rise to a favorable reaction fieldlike ligand–protein interaction. It might be possible to take both these effects into account within the framework of a quasiharmonic representation of protein fluctuations.<sup>47,48</sup> In the present effort, we will attempt to incorporate the dominant effects by assigning a nonunity dielectric constant  $\epsilon_{\text{prot}}$  to the outer region protein interior. At the crudest level, this simple approximation may be able to capture the essential aspects of the electrostatic contributions arising from the outer region, namely, the shielding of the static field and the induced reaction field. The ability of a protein dielectric constant to represent the physical characteristics of protein fluctuations is still a widely debated issue<sup>49–52</sup> and a more sophisticated treatment might be needed in the future.

To examine the influence of the protein dielectric constant, FEP calculations were performed with values of  $\epsilon_{\text{prot}}$  varying from 1 to 4 for all three simulated systems. The protein static field  $\phi_s^{(0)}(\mathbf{r})$  and the reaction field matrix  $\mathbf{M}$  were recalculated using these values of  $\epsilon_{\text{prot}}$  (note that in all cases the dielectric constant of the inner region was set to 1). The results of the FEP calculations are given in Table III. In these calculations, the protein atoms in the inner region were allowed to move while only the aspartate ligand was held fixed in space. The results of the FEP calculations show that the charging free energy is not greatly influenced by the value of  $\epsilon_{\text{prot}}$  in the case of the 20 Å sphere size. The charging free energy of the aspartate ligand in a fixed configuration in the active site of aspartyl-tRNA synthetase appears

to converge towards a value of approximately 109 to 110 kcal/mol, which we use as a reference to assess the other calculations. In the case of the 15 Å sphere size, the approximate treatment improves the results and an optimal value is reached when  $\epsilon_{\text{prot}}$  is between 3 and 4, while the calculation does not converge to a correct value in the case of the 10 Å sphere size. It is to be noted that the latter comprises only 468 atoms (416 are mobile), and that the nearest fixed protein atom is only 7.3 Å away from the aspartate ligand. Such a system is probably too small and failure of the approximation of a uniform protein dielectric constant should not be unexpected (see below).

As seen from Table III, the best estimate for the charging free energy in the flexible protein is very similar to the charging free energy in liquid water given in Table II. This suggests that the total electrostatic interactions in the two system are roughly counterbalanced. Although explicit experimental data on the equilibrium binding constant of aspartic acid to aspartyl-tRNA synthetase could not be found in the literature, the estimated  $K_m$  values for aspartic acid in the aminoacylation reaction are in the micromolar range,<sup>53</sup> indicating that a rough estimate of the binding free energy is around  $-6$  kcal/mol. In comparison to results from Table I, it is clear that the calculated charging free energy with fixed protein is roughly 60 kcal/mol more favorable. This highlights the importance of protein flexibility in estimating binding free energies. Ignoring the outer region interactions altogether by considering a truncated system results in grossly incorrect values for all sphere sizes, with errors of 74, 34, and 22 kcal/mol for the 10, 15, and 20 Å sphere size, respectively. This result also reinforces the importance of a correct treatment of the long-range electrostatic interactions in such FEP calculations. Interestingly, approximately correct results are obtained for the 15 and 20 Å sphere size, even when the static field is neglected as long as the reaction field is included. This suggests that, at least in the case of aspartyl-tRNA synthetase, the static field from protein regions beyond 15 to 20 Å from the ligand is significantly shielded. In contrast, ignoring the reaction field yields errors on the order of 33 and 26 kcal/mol, for the 15 and 20 Å sphere size, respectively. These observations provide further support for the computational scheme used by Simonson in which the static field is neglected.<sup>33</sup>

The most severe approximation in the present hybrid MD/continuum approach is that the protein atoms beyond the inner region are fixed. For example, significant errors in

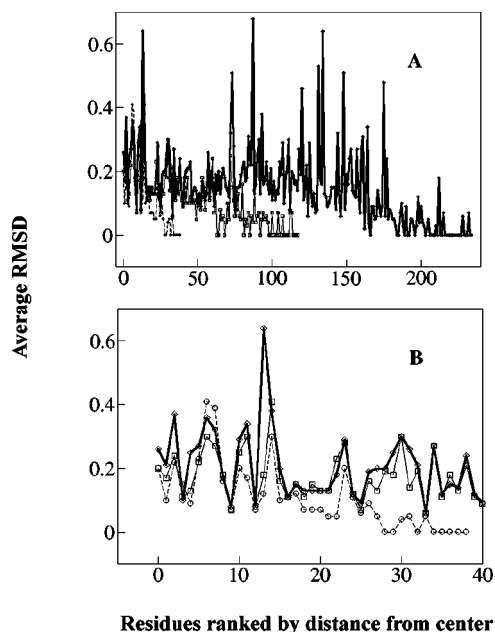


FIG. 1. The average root mean square difference (RMSD) values of individual protein residues during the FEP simulations. 10 Å inner region sphere: dotted line with circles; 15 Å inner region sphere: thin line with squares; 20 Å inner region sphere: bold line with diamonds. Residues are ranked based on their presence in 1 Å radial distance increments from the center of the sphere and are plotted in increasing order of rank. (a) The average RMSD for inner region residues in all three sphere sizes; (b) the average RMSD for the first 40 residues (roughly corresponding to a distance of less than 10 Å from the center of the sphere) for the three sphere sizes. All average RMSD values obtained for nonhydrogen atoms from simulations with an outer protein dielectric constant set to 1.0; sampling done every 1 ps.

the results can arise from protein residues beyond the flexible inner region that would reorient spontaneously when the ligand changes or disappears. To identify the presence of such residues, the average values of the root mean square difference (RMSD) of nonhydrogen atoms belonging to each residue were calculated in comparison to the starting structure. In Fig. 1, these values are plotted as a function of their distance from the common center of the inner region spheres. It is observed that some residues beyond the smaller inner region spheres show greater flexibility than residues in the inner region. This is true both for the 10 Å and 15 Å inner spheres and indicates a fundamental caveat in hybrid MD/continuum approaches. About 36 percent of the residues that show greater than 1.0 Å RMSD at any point during the FEP simulations in the 20 Å sphere size are distant negatively charged residues that lie in the same large solvent exposed cavity as the aspartate ligand. Can a uniform dielectric constant assigned to the protein in the outer region approximate the effect of their complex motions? The results with the 10 Å sphere size indicate that a simple increase in  $\epsilon_{\text{prot}}$  is unable to capture the complexity of the electrostatic response of a heterogeneous protein if the flexible inner region is too small. In this case, increasing  $\epsilon_{\text{prot}}$  has the disappointing effect of increasing the calculated magnitude of the charging free energy. This indicates that the ligand is more favorably interacting with its protein-solvent environment. The reason for this trend is clear from Table I. Both static field and

TABLE IV. Charging free energy PB calculations in the fixed ligand-protein complex and with changing protein conformation. Calculations performed using the PBEQ module in CHARMM (Ref. 35) and values reported in kcal/mol. The value in parenthesis is the standard deviation for 25 different conformations of protein around fixed ligand.

Protein dielectric constant	Reentrant, probe radius 1.4 Å	Reentrant, variable protein conformation	Nonreentrant
1.0	-161.3	-163.5 (7.4)	-161.3
2.0	-120.2	-	-122.1
3.0	-102.8	-	-104.0
4.0	-93.0	-	-94.1

reaction field electrostatic interactions of the outer region with the ligand for the 10 Å sphere size system are repulsive and a significant increase in the calculated charging free energy is caused when they are completely ignored. An increase in  $\epsilon_{\text{prot}}$  which results in a shielding of these repulsive contributions from the outer region also has a similar, albeit less pronounced, effect, thereby making the answer worse than when  $\epsilon_{\text{prot}}$  is set to 1. In the case of the 15 Å sphere size, the change in the calculated free energy due to a change in  $\epsilon_{\text{prot}}$  is almost entirely due to the change in the static field component, while the static field and reaction field contribution changes caused by a change in  $\epsilon_{\text{prot}}$  are balanced by each other for the 20 Å sphere size, such that the overall charging free energy seems to be insensitive to the value of  $\epsilon_{\text{prot}}$ . Thus it may be concluded that the choice of both inner region system size and dielectric constant for the outer region protein must be considered carefully in the application of GSBP.

It is of interest to compare the present results to those obtained with a continuum electrostatic approximation using a standard application of the PB equation.<sup>23,54</sup> The total charging free energy of the ligand (L) bound to the protein (P) is calculated as

$$\Delta G = G_{\text{PL}}(\epsilon_{\text{solvent}} = 80) - G_{\text{P}}(\epsilon_{\text{solvent}} = 80) - G_{\text{L}}(\epsilon_{\text{solvent}} = 1), \quad (8)$$

where the  $G$ 's represent the charging free energy of each particular configuration. The charging free energy estimations of the aspartic acid-aspartyl-tRNA complex were calculated for a fixed conformation of the equilibrated complex. The results are given in Table IV. The average charging free energy for different conformations of the protein around the fixed ligand is also reported for the reentrant surface and protein dielectric constant of 1. The close correspondence of this average value and the value obtained for a single conformation of the protein illustrates that the chosen equilibrated protein conformation used in all other PBEQ calculations is a good representative of the average. The method based on Eq. (8) implicitly assumes that a single value for the dielectric constant of the solutes (protein and ligand) is applied equivalently to both the static and reaction field components of the charging free energy. Modifications to this method, in which different protein dielectric constants are used for the static field and the reaction field components, have also been utilized.<sup>23,25</sup> The  $\Delta G$  obtained using a protein dielectric constant of 1 is 161 kcal/mol, which is quite close

TABLE V. Effect of molecular surface description on charging free energy PB calculations in the fixed ligand–protein complex with protein charges removed. Calculations performed using the PBEQ module in CHARMM (Ref. 35) and values reported in kcal/mol. Protein dielectric set to 1 and solvent dielectric set to 80.

Solvent probe radius	Charging free energy
0.0	−86.7
0.1	−82.7
0.2	−75.9
0.3	−69.8
0.4	−64.9
0.5	−60.2
0.6	−56.5
0.7	−54.0
0.8	−51.6
0.9	−47.1
1.0	−41.7
1.1	−37.3
1.2	−35.1
1.3	−33.5
1.4	−31.4
1.5	−29.0

to the corresponding value of 172 kcal/mol obtained using FEP/GSBP. Increasing the protein dielectric constant results in a decrease in the calculated free energy, which is consistent with the corresponding decrease observed in the FEP/GSBP results with flexible inner region. The solute–solvent boundary for a PB calculation can either be obtained by the van der Waals surface (overlapping atomic spheres) or as a solvent-accessible molecular surface (including reentrants) with a probe of 1.4 Å radius. For the van der Waals surface, the results with a protein dielectric constant of 1 remain unchanged (161 kcal/mol) as compared to the solvent-accessible molecule surface. The results obtained thus appear to be independent of the exact placement of the solute–solvent boundary.

The dielectric boundary, however, is expected to affect both the static field and reaction field components of the charging free energy. The absence of a significant difference between nonreentrant and reentrant surfaces may be due to a balance between changes in the two components rather than both components remaining unaffected. To test the influence of the solute–solvent boundary on the reaction field component alone, the charging free energy values with all protein charges turned off (no protein static field) were determined by solving the PB equation. The results are given in Table V. Increasing the solvent probe radius from 0.0 Å (nonreentrant surface) to 1.5 Å is seen to cause significant variation in the calculated charging free energy. The best agreement with the FEP/GSBP results for the largest sphere (−71 kcal/mol) is obtained with a solvent probe radius of 0.4 Å. The commonly used solvent probe radius of 1.4 Å gives a much smaller magnitude for the charging free energy (−31 kcal/mol). Not having the proper solvent probe radius can lead to artificial high dielectric regions in the interior of the protein. In this case, however, the lack of protein static field means that the variation of the results seen above is most likely due to the presence or absence of high dielectric solvent regions in close proximity to the charged ligand rather

than remote regions of the protein. Use of a hybrid continuum/MD method such as FEP/GSBP can avoid this problem by having an explicit solvent description in the vicinity of the charged ligand. Alternatively, it is possible to choose a protein dielectric constant in conjunction with an appropriate description of the molecular surface and probe radius to accurately reproduce charging free energies using the PB equation. The caveats are that the optimal value of the protein dielectric constant is highly system-dependent<sup>52</sup> and the charging free energy calculated is also sensitive to small inaccuracies in the solute–solvent boundary. Since these difficulties are circumvented in a hybrid MD/continuum approach, the use of more detailed methods may be necessary to validate the parameters used in the PB equation before it can be applied with reasonable confidence for the determination of charging free energies for a novel ligand–protein complex.<sup>24</sup>

#### IV. CONCLUSION

In the present study we show that the GSBP method can allow an efficient treatment of the long-range interactions by using explicit solvent MD in a small inner region and treatment of the outer region using an implicit solvent model. Initial tests with all protein and solute atoms held fixed were used to examine the consistency of GSBP. The small variations with system size in the final charging free energy of aspartate (less than 3 percent) demonstrates that the solvent-shielded static field and solvent-induced reaction field of GSBP correctly represent the implicit solvent region. Importantly, ignoring the reaction field contribution results in gross errors in the calculated charging free energy.

The dominant effects of protein flexibility were incorporated by assigning a nonunity dielectric constant to the outer region protein interior. Not surprisingly, the magnitude of the charging free energy is significantly affected by the flexibility of the inner region protein atoms. This highlights the importance of protein fluctuations in the determination of ligand binding free energy. Although this approximation failed in the case of the smallest system of 10 Å, consistent results were obtained with system size of 15 and 20 Å using a protein dielectric constant of 3 or 4. Since those two simulation systems comprises roughly 1500 and 3500 atoms, respectively, accurate results can thus be obtained using relatively small simulation systems. In the future, it might be possible to push this limit further with a more sophisticated treatment of protein fluctuations in the outer region.

It is difficult to assess the convergence of the present series of GSBP calculations in the absence of an absolute free energy reference. A massive, traditional FEP simulation with the entire protein solvated by a large number of explicit water molecules (roughly 160,000 atoms) with periodic boundary conditions could, in principle, provide such a reference. However, such a calculation would face significant difficulties ranging from sampling and convergence issues, to the treatment of long-range electrostatics,<sup>55</sup> and is not feasible at the present time. However, despite the lack of an exact answer to compare the GSBP results with, the internal consistency observed in the series of calculations shown here is very satisfying given the fact that, for each system size, the



contribution from the explicit atoms, the protein static field, and the reaction field (represented by a generalized basis set expansion) are completely different.

The approximation of the fixed ligand, though useful for the purposes of a systematic comparison in the present study, limits any comparison with experiments. A realistic estimation of the binding constant would have to take into account the mobility of the ligand, and various factors corresponding to translational and rotational entropy.<sup>56,57</sup> Irrespective of these considerations, the present calculations show that easy, accurate, and meaningful FEP estimations are possible using the FEP/GSBP approach. It is our hope that GSBP can become a useful and efficient computational tool to screen large molecular databases and identify potential lead drug molecules.<sup>7</sup>

## ACKNOWLEDGMENTS

The authors would like to thank Simon Bernèche and Guillaume Lamoureux for helpful discussions regarding the FEP calculations. This work was supported by National Science Foundation (NSF) Grant No. 529110 and National Institutes of Health (NIH) Grant No. 525408.

- <sup>1</sup>A. Vindigni, *Comb. Chem. High Throughput Screening* **2**, 139 (1999).
- <sup>2</sup>A. C. Cheng, V. Calabro, and A. D. Frankel, *Curr. Opin. Struct. Biol.* **11**, 478 (2001).
- <sup>3</sup>C. W. Garvie and C. Wolberger, *Mol. Cell.* **8**, 937 (2001).
- <sup>4</sup>J. B. Chaires, *Anti-Cancer Drug Des.* **11**, 569 (1996).
- <sup>5</sup>W. H. Ward and G. A. Holdgate, *Prog. Med. Chem.* **38**, 309 (2001).
- <sup>6</sup>A. Wlodawer, *Annu. Rev. Med.* **53**, 595 (2002) (in English).
- <sup>7</sup>G. Schneider and H. J. Bohm, *Drug Discov. Today* **7**, 64 (2002).
- <sup>8</sup>N. D. Lu and D. A. Kofke, *J. Chem. Phys.* **114**, 7303 (2001).
- <sup>9</sup>N. D. Lu and D. A. Kofke, *J. Chem. Phys.* **115**, 6866 (2001).
- <sup>10</sup>J. Florian, M. F. Goodman, and A. Warshel, *J. Phys. Chem. B* **104**, 10092 (2000).
- <sup>11</sup>B. O. Brandsdal and A. O. Smalas, *Protein Eng.* **13**, 239 (2000).
- <sup>12</sup>P. Kollman, *Chem. Rev.* **93**, 2395 (1993).
- <sup>13</sup>D. L. Beveridge and F. M. Dicapua, *Annu. Rev. Biophys. Bioeng.* **18**, 431 (1989).
- <sup>14</sup>C. Chipot, P. A. Kollman, and D. A. Pearlman, *J. Comput. Chem.* **17**, 1112 (1996).
- <sup>15</sup>T. Simonson, G. Archontis, and M. Karplus, *J. Phys. Chem. B* **101**, 8349 (1997).
- <sup>16</sup>C. Sagui and T. A. Darden, *Annu. Rev. Biophys. Biomol. Struct.* **28**, 155 (1999).
- <sup>17</sup>M. Berkowitz and J. McCammon, *Chem. Phys. Lett.* **90**, 215 (1982).
- <sup>18</sup>C. L. Brooks III and M. Karplus, *J. Chem. Phys.* **79**, 6312 (1983).
- <sup>19</sup>A. T. Brunger, C. L. Brooks III, and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8458 (1985).
- <sup>20</sup>C. L. Brooks III and M. Karplus, *J. Mol. Biol.* **208**, 159 (1989).
- <sup>21</sup>A. Warshel and A. Papazyan, *Curr. Opin. Struct. Biol.* **8**, 211 (1998).
- <sup>22</sup>G. Archontis, T. Simonson, D. Moras, and M. Karplus, *J. Mol. Biol.* **275**, 823 (1998).
- <sup>23</sup>T. Simonson, G. Archontis, and M. Karplus, *J. Phys. Chem. B* **103**, 6142 (1999).
- <sup>24</sup>G. Archontis, T. Simonson, and M. Karplus, *J. Mol. Biol.* **306**, 307 (2001).
- <sup>25</sup>G. Archontis and T. Simonson, *J. Am. Chem. Soc.* **123**, 11047 (2001).
- <sup>26</sup>W. Im, S. Bernèche, and B. Roux, *J. Chem. Phys.* **114**, 2924 (2001).
- <sup>27</sup>D. Beglov and B. Roux, *J. Chem. Phys.* **100**, 9050 (1994).
- <sup>28</sup>J. G. Arnez and D. Moras, *Trends Biochem. Sci.* **22**, 211 (1997).
- <sup>29</sup>C. Berthet-Colominas, M. H. L. Seignovet, M. Grotli, S. Cusack, and R. Leberman, *EMBO J.* **17**, 2947 (1998).
- <sup>30</sup>S. Eiler, A. Dock-Bregeon, L. Moulinier, J. C. Thierry, and D. Moras, *EMBO J.* **18**, 6532 (1999).
- <sup>31</sup>C. Briand, A. Poterszman, S. Eiler, J. T. G. Webster, and D. Moras, *J. Mol. Biol.* **299**, 1051 (2000).
- <sup>32</sup>L. Moulinier, S. Eiler, G. Eriani, J. Gangloff, J. C. Thierry, K. Gabriel, W. H. McClain, and D. Moras, *EMBO J.* **20**, 5290 (2001).
- <sup>33</sup>T. Simonson, *J. Phys. Chem. B* **104**, 6509 (2000).
- <sup>34</sup>M. Nina, D. Beglov, and B. Roux, *J. Phys. Chem. B* **101**, 5239 (1997).
- <sup>35</sup>W. Im, D. Beglov, and B. Roux, *Comput. Phys. Commun.* **111**, 59 (1998).
- <sup>36</sup>B. Roux, *Biophys. J.* **73**, 2980 (1997).
- <sup>37</sup>B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- <sup>38</sup>I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig, *Proteins* **1**, 47 (1986).
- <sup>39</sup>A. Nicholls and B. Honig, *J. Comput. Chem.* **12**, 435 (1991).
- <sup>40</sup>M. K. Gilson, K. Sharp, and B. Honig, *J. Comput. Chem.* **9**, 327 (1987).
- <sup>41</sup>C. L. Brooks III and M. Karplus, *J. Mol. Biol.* **208**, 159 (1989).
- <sup>42</sup>J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- <sup>43</sup>R. H. Stote, D. J. States, and M. Karplus, *J. Chim. Phys. Phys.-Chim. Biol.* **88**, 2419 (1991).
- <sup>44</sup>S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, *J. Comput. Chem.* **13**, 1011 (1992).
- <sup>45</sup>M. Souaille and B. Roux, *Comput. Phys. Commun.* **135**, 40 (2001).
- <sup>46</sup>L. Banci, S. Schroder, and P. A. Kollman, *Proteins* **13**, 288 (1992).
- <sup>47</sup>X. Song, *J. Chem. Phys.* **116**, 9359 (2002).
- <sup>48</sup>T. Simonson and D. Perahia, *Comput. Phys. Commun.* **91**, 291 (1995).
- <sup>49</sup>T. Simonson and C. L. Brooks III, *J. Am. Chem. Soc.* **118**, 8452 (1996).
- <sup>50</sup>Y. Y. Shan, I. Muegge, and A. Warshel, *Biophys. J.* **74**, 1744 (1998).
- <sup>51</sup>J. J. Dwyer, A. G. Gittis, D. A. Karp, E. E. Lattman, D. S. Spencer, W. E. Stites, and E. B. Garcia-Moreno, *Biophys. J.* **79**, 1610 (2000).
- <sup>52</sup>C. N. Schutz and A. Warshel, *Proteins* **44**, 400 (2001).
- <sup>53</sup>F. Martin, G. J. Sharples, R. G. Lloyd, S. Eiler, D. Moras, J. Gangloff, and G. Eriani, *J. Bacteriol.* **179**, 3691 (1997).
- <sup>54</sup>V. K. Misra, J. L. Hecht, A. S. Yang, and B. Honig, *Biophys. J.* **75**, 2262 (1998).
- <sup>55</sup>P. H. Hunenberger and J. A. McCammon, *Biophys. Chem.* **78**, 69 (1999).
- <sup>56</sup>B. Roux, M. Nina, R. Pomes, and J. C. Smith, *Biophys. J.* **71**, 670 (1996).
- <sup>57</sup>M. Gilson, J. Given, B. Bush, and J. McCammon, *Biophys. J.* **72**, 1047 (1997).