Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the POLARIS and ENZYMIX Programs

Frederick S. Lee, Zhen Tao Chu, and Arieh Warshel*

Department of Chemistry, University of Southern California, Los Angeles, California 90089-1062

Received 21 April 1992; accepted 4 August 1992

Different microscopic and semimicroscopic approaches for calculations of electrostatic energies in macromolecules are examined. This includes the Protein Dipoles Langevin Dipoles (PDLD) method, the semimicroscopic PDLD (PDLD/S) method, and a free energy perturbation (FEP) method. The incorporation of these approaches in the POLARIS and ENZYMIX modules of the MOLARIS package is described in detail. The PDLD electrostatic calculations are augmented by estimates of the relevant hydrophobic and steric contributions, as well as the effects of the ionic strength and external pH. Determination of the hydrophobic energy involves an approach that considers the modification of the effective surface area of the solute by local field effects. The steric contributions are analyzed in terms of the corresponding reorganization energies. Ionic strength effects are studied by modeling the ionic environment around the given system using a grid of residual charges and evaluating the relevant interaction using Coulomb's law with the dielectric constant of water. The performance of the FEP calculations is significantly enhanced by using special boundary conditions and evaluating the long-range electrostatic contributions using the Local Reaction Field (LRF) model. A diverse set of electrostatic effects are examined, including the solvation energies of charges in proteins and solutions, energetics of ion pairs in proteins and solutions, interaction between surface charges in proteins, and effect of ionic strength on such interactions, as well as electrostatic contributions to binding and catalysis in solvated proteins. Encouraging results are obtained by the microscopic and semimicroscopic approaches and the problems associated with some macroscopic models are illustrated. The PDLD and PDLD/S methods appear to be much faster than the FEP approach and still give reasonable results. In particular, the speed and simplicity of the PDLD/S method make it an effective strategy for calculations of electrostatic free energies in interactive docking studies. Nevertheless, comparing the results of the three approaches can provide a useful estimate of the accuracy of the calculated energies.

INTRODUCTION

Electrostatic energies provide what is probably the most powerful correlation of structure and function in biologic molecules.[1-6] However, obtaining stable and reliable results from electrostatic calculations of proteins and other macromolecules is far from being simple due to the long-range nature of electrostatic effects and the nonhomogeneous nature of the microenvironments of biologic molecules.[2-4] Microscopic calculations of electrostatic energies encounter major problems in terms of their accuracies because they reflect the compensation of large energy contributions and frequently involve a rather small cutoff radius for dipolar interactions. Macroscopic calculations, on the other hand, involve the use of *assumed* dielectric constants that cannot be deduced from macroscopic considerations without direct experimental information about the energetics of the relevant groups in the given macromolecule.[4]

Probably the best way to select an optimal approach for electrostatic calculations of macromolecules is to examine the ability of different alternative models to reproduce the observed energetics of many different test cases. In fact, many of the traps encountered by early macroscopic studies (see the discussion in ref. 7) could have been avoided if these methods were used to calculate other properties in addition to charge–charge interactions between surface groups.

This work adopts the above philosophy and considers an extensive set of test cases ranging from the solvation of ions in water, self-energies of charges in proteins, energetics of ion pairs in proteins, and electrostatic contributions to catalysis and binding. The methods examined include an adiabatic charging FEP approach that is rigorous in principle (with infinite number of solvent molecules and infinite simulation time), the simplified microscopic model of the PDLD method, and the recently intro-

*Author to whom all correspondence should be addressed.

CCC 0192-8651/93/020161-25

duced PDLD/S method. In addition, we examine the performance of some commonly used macroscopic models and demonstrate that the energetics obtained by such models can be quite inaccurate. The comparison of the different methods provides useful information about the range of applicability and speed of each of these approaches.

Significant attention is given to the performance of the PDLD/S method. This method has been introduced to avoid some of the difficulties associated with the precision of the microscopic models by exploiting the inherent precision of the macroscopic models, that is, the macroscopic models are more precise than the microscopic models because the electrostatic potentials and fields are scaled by a dielectric constant that represents the assumed compensation between the different microscopic contributions. However, the macroscopic models can be less accurate than the microscopic models when applied to charges in the interiors of proteins (where the proper dielectric constant is not known). Fortunately, it is possible to use the PDLD/S method to *scale* the microscopic PDLD contributions in a way that avoids the main pitfalls associated with most early macroscopic studies of proteins and obtain an expression that considers explicitly the protein permanent dipoles but scales their effects and the relevant self-energies by an assumed dielectric constant, ϵ_{in}^{a} (see refs. 4 and 8 for the actual meaning of this "dielectric constant"). It is found that the PDLD/S method can provide useful estimates of electrostatic energies in macromolecules. However, a combined use of the PDLD, PDLD/S, and FEP methods is highly recommended because it can provide a clear indication of the error range associated with the estimate of the given electrostatic energy.

THEORETICAL METHODS

This section considers our microscopic approaches for calculations of electrostatic energies and their implementations in the corresponding simulation programs. The problems and relative advantages associated with each approach will be considered in the results section.

Implementation of the Simplified Microscopic PDLD Method in POLARIS

General Considerations

The PDLD model emerged early[2-4,9] as a *practical* way of evaluating electrostatic energies in proteins on a microscopic level. This was essential at that stage to avoid the uncertainties associated with the use of macroscopic models in macromolecules. The PDLD method approximates the *average* energy of the protein and water system by the electrostatic energy evaluated at the average structure. The average protein structure can be taken from the X-ray structure (corresponding to a time-average in the crystalline state), which is sometimes reminimized for different charge configurations. It is also possible to average the calculations over different structures generated by molecular dynamics (MD) simulations (see below). The average polarization of the solvent molecules is represented as follows (see ref. 3 for more details and rationales): The protein is surrounded by a three-dimensional cubic grid and each grid point that is within a specific van der Waals distance from a protein atom is deleted. The grid is truncated to a sphere and each of its remaining points is occupied by a point dipole that represents the average polarization a water molecule at that site. Each point dipole is allowed to be polarized toward the local field due to the protein atoms as well as other solvent dipoles *except* the nearest neighbors (see ref. 3). Thus, the local field on the ith Langevin dipole can be expressed as

$$\boldsymbol{\xi}_i = \boldsymbol{\xi}_i^0 + \boldsymbol{\xi}_i^L - \boldsymbol{\xi}_i^c \qquad (1)$$

where $\boldsymbol{\xi}^0$ is the field due to the protein permanent charges (evaluated only once because the protein structure is fixed), $\boldsymbol{\xi}^L$ is the field due to *all* other Langevin dipoles, and $\boldsymbol{\xi}^c$ is the field due to the Langevin dipoles that are the nearest neighbors to the ith dipole.* The polarization of the ith Langevin dipole by its local field $\boldsymbol{\xi}_i$ of eq. (1) is approximated by a Langevin-type function[3]:

$$(\boldsymbol{\mu}_i^L)^{n+1} = \mathbf{e}_i^n \mu_0 \left(\coth \chi_i^n - \frac{1}{\chi_i^n} \right) \qquad (2)$$

$$\chi_i^n = \frac{C' \mu_0}{k_B T} |\boldsymbol{\xi}_i^n|$$

where $\mathbf{e}_i$ is a unit vector in the direction of the local field $\boldsymbol{\xi}_i$, C' is a parameter, and μ_0 is taken as 1.8 D. The equation for the effective Langevin dipoles, $\boldsymbol{\mu}_i^L$, is solved iteratively ($\boldsymbol{\mu}_i^{n+1}$ is determined by the field $\boldsymbol{\xi}_i^n$ from the previous iteration). The relevant polarization law can be obtained by calibrating the parameter C' and the solute–solvent van der Waals' distances for different atom types so that the solvation energies of various ionic species with different charges and radii are reproduced. This rather

*The field $\boldsymbol{\xi}_i$ used in ref. 9 was replaced in some PDLD versions (e.g., ref. 10) by $\boldsymbol{\xi}_i^0$, giving faster convergence and results similar to those obtained with $\boldsymbol{\xi}_i^0 + \boldsymbol{\xi}_i^L$. This useful treatment, that was recently criticized,[11] is a reasonable approximation for models that represent permanent dipoles by their projections, as demonstrated in Figure 5 of ref. 3. The key idea here is that the distribution function for the average polarization of the solvent follows some given polarization law and a model that reproduces this law also would reproduce the electrostatic interaction between the solute and the solvent. Of course, such model does not have to reproduce the exact structure of the solvent or the exact behavior of the corresponding dipolar molecule. It should only reproduce the proper polarization.

simple solvent model appears to give quite reliable solvation energies, reflecting the fact that the physics of solvation effects can be described by many alternative dipolar models,[2-4] that is, the key for obtaining reliable energetics lies in the consistent calibration of the solvent residual charges and nonbonded parameters using observed solvation energies and radial distribution functions (because the residual charges are highly correlated with the other nonbonded parameters, it is possible to get similar results with different parameter sets). The above statement might sound strange to some because it is frequently stated in the literature that the hydrogen bonding properties of different solvents are special. However, although different solvents might have different structural properties it appears that different polar solvents (with and without hydrogen bonds) give similar solvation energies. For instance, the free energy of solvation for a Na^+ ion in water and in ammonia is -98 and -96 kcal mol^{-1}, respectively. Apparently, the physics of electrostatic interactions involves many compensating-effects so that the overall solvation energy turns out to be similar even though its individual components are quite different. This might be one reason why solvation effects can be reproduced even with rather simple continuum models using an adjustable cavity radius.[12] It might be useful to point out here that some workers who found macroscopic models completely valid have felt that the PDLD model represents a major approximation.[13] Apparently, it is hard to appreciate the difference between the approximations involved in the simplified microscopic models (such as the PDLD model) and those involved in macroscopic models. In fact, the PDLD model approximates the solvent by dipoles and then approximates the average polarization of each dipole by a Langevin dipole, while continuum models approximate a collection of dipoles by a volume element with a uniform polarization (see Fig. 1), which is a much more serious approximation. It is also useful to clarify that the PDLD model does not represent a static polarization[13] but the time average polarization of the solvent (which reflects the full dynamics of the system). In fact, it is instructive to comment that a recent Langevin dynamics simulation of point dipoles on a grid reproduced the Langevin polarization[14] (see also Fig. 5 of ref. 3).

The PDLD model also explicitly includes the effect of the protein electronic polarization by assigning induced dipole moments to all the protein atoms. These dipoles interact with the permanent charge distribution of the system and the Langevin dipoles, as well as with each other, and are therefore treated by an iterative procedure[9,10] (although an effective noniterative approach is also available[9,10]).

The present version of the PDLD model, implemented in the program POLARIS,[15] divides the protein into three regions as depicted in Figure 2. Region

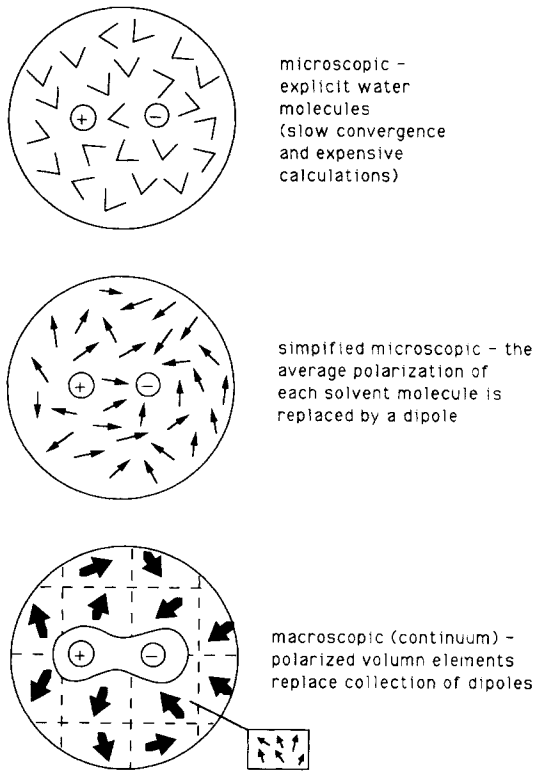


Figure 1. Relationship among the microscopic, simplified microscopic, and macroscopic models. The microscopic model uses detailed all-atom representation and evaluates the average interaction between the solvent residual charges and the solute charges. Such calculations are expensive. The simplified microscopic model replaces the time average dipole of each solvent molecule by a point dipole, while the macroscopic model considers the collection of solvent dipoles in a large volume element as a polarization vector.

1 contains the "solute" or the groups whose electrostatic energy is of interest. Region 2 contains the charged and electroneutral groups within a radius R_2 from the center of region 1. The protein groups beyond R_2 are excluded from the calculations and the corresponding space (region 4a) is occupied by the solvent (in the representation of either region 3 or 4). This spherical truncation makes the electrostatic calculations more stable. The solvent region, referred to as region 3, is modeled by the abovementioned spherical grid of Langevin dipoles of a radius R_3. The grid has two spacings: The inner region (up to a distance R_{in} from the closest protein atom) has a spacing of 2 Å whereas the remaining volume has a spacing of 3 Å. The dipole moment, $\boldsymbol{\mu}^L$ of eq. (2), at any grid point is scaled by the factor $(\Delta/3)^3$ where Δ is the grid spacing of the given region. The rationale for a smaller inner grid spacing for the dipoles surrounding region 1 is to increase the stabilities of the calculations by giving smaller energy differences for different grid configurations. Grid points that are found within the van der Waals radius of any protein or ligand atom are deleted. The solvent region beyond R_3 (referred to as region 4) is

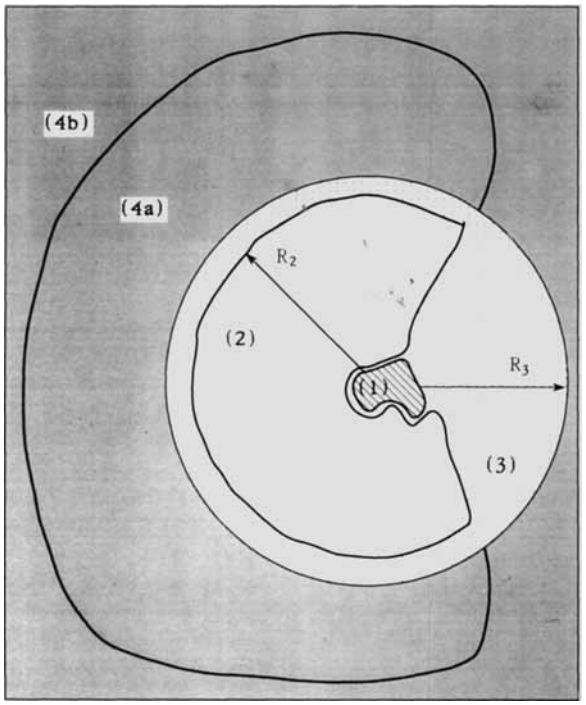


Figure 2. Four regions of the protein system in the PDLD method. Region 1 contains the charged groups of interest. Region 2 contains the protein atoms found within a radius R_2 from the center. Region 3 is the Langevin grid truncated to a sphere of a radius R_3. Region 4a contains the rest of the protein atoms outside region 2. The electrostatic effects of regions 1, 2, and 3 are treated explicitly while those of region 4 (4a and 4b) are considered as bulk solvent regions and treated by a macroscopic continuum formulation. Note that the protein in region 4a is replaced by bulk solvent.

represented by a continuum model with the dielectric constant of water ($\epsilon = \epsilon_w$).

The electrostatic energy of the PDLD model is evaluated by considering the different contributions to the energy of the solute in the surrounding environment (regions 1, 2, 3, and 4). These contributions include the following:

1. The interaction between the charges in the solute region (region 1), given by

$$V_{QQ} = 332 \sum_{i<j} \frac{Q_i Q_j}{r_{ij}} \tag{3}$$

(throughout this work, energies are expressed in kcal mol^{-1}, charges in units of electron charge, and distances in Å).

2. The interaction between the charges of region 1 and the residual charges of the protein (the charges of region 2), given by

$$V_{Q\mu} = 332 \sum_{ij} \frac{Q_i q_j}{r_{ij}} \tag{4}$$

Here, we denote the charges of region 1 and 2 by Q and q, respectively.

3. The energy of the protein-induced dipoles, which is given by[3]

$$V_{Q\alpha} = -\frac{332}{2} \sum_i \boldsymbol{\mu}_i^p \boldsymbol{\xi}_i^o \tag{5}$$

where $\boldsymbol{\xi}_i^o$ is the field on the ith protein atom from the protein permanent charges in regions 1 and 2 and $\boldsymbol{\mu}_i^P$ is the protein-induced dipoles associated with the polarization of the electrons of the ith protein atom. This dipole is given by

$$\boldsymbol{\mu}_i^p = \alpha_i \boldsymbol{\xi}_i \tag{6}$$

where α_i is the atomic polarizability of the ith protein atom and $\boldsymbol{\xi}_i$ is the local field on the ith atom due to $\boldsymbol{\xi}_i^o$, the other protein-induced dipoles and optionally the Langevin dipoles. In this model, the solute atoms in region 1 are not polarizable (i.e., their αs are set to zero). The polarization of the solute atoms can be evaluated by treating it quantum mechanically.[16] The factor ½ in eq. (5) reflects the fact that half the energy associated with the dipole–field interaction is invested in forming the induced dipole.[3] The field $\boldsymbol{\xi}_i^o$ is evaluated with a screening function, f_d [which prevents overpolarization at close distances where the classic approximation of eq. (6) is invalid[3]], and is given by

$$\boldsymbol{\xi}_i^o = \sum_j \frac{q_j \mathbf{r}_{ij} f_d(r_{ij})}{r_{ij}^3} \tag{7}$$

$$f_d(r_{ij}) = 1 - \exp(-r_{ij}^4/r_d^4)$$

where r_d is taken as 3 Å.

4. The solvation energy of the solute and the protein by the surrounding water regions (regions 3 and 4) is given by[3]

$$\Delta G_{Qw} = \Delta G_{lgvn} + \Delta G_{bulk} \tag{8}$$

$$\Delta G_{lgvn} = -\frac{332}{2} \sum_k \boldsymbol{\mu}_k^L \boldsymbol{\xi}_k^o$$

$$\Delta G_{bulk} = -\frac{166 Q_b^2}{R_{lgv}} \left(1 - \frac{1}{\epsilon_w}\right) - \frac{166 \mu_b^2}{R_{lgv}^3} \frac{(2\epsilon_w - 2)}{(2\epsilon_w + 1)}$$

where ΔG_{lgvn} the energy of Langevin dipoles (in region 3) and ΔG_{bulk} is the electrostatic energy associated with the interaction between our finite grid system and its infinite surrounding (region 4), which is represented as a continuum. $\boldsymbol{\mu}_k^L$ is evaluated by eq. (2) and $\boldsymbol{\xi}_k^o$ is the field on the kth Langevin dipole from the protein permanent charge distribution. The factor ½ reflects the fact that half the energy gained from the dipole–charge interaction is invested in polarizing the solvent dipoles.[3] Q_b and μ_b are the monopole and dipole moments of the solute at the designated center of the system. ϵ_w is the dielectric constant of water, taken as 80 throughout this work.

Local Reaction Field Corrections

One of the most time-consuming steps in the PDLD calculation is the evaluation of the field ξ^L of eq. (1) at the site of each Langevin dipole. Such calculations require one to consider the contributions to ξ_i^L from Langevin dipoles quite far from the ith dipole. These calculations can be accelerated by using a small cutoff radius and truncating the long-range interaction of the Langevin dipoles. This approach, however, leads to disappointing results. The long-range problem can be partially overcome by the Local Reaction Field (LRF) method.[17] The LRF method calculates explicitly the short-range electrostatic potential at the site of each group by a standard truncation procedure (considering all the contributions from within the given cutoff radius), while approximating the long-range contributions by a fourth-order expansion that is updated at a much slower rate than the time steps used for the given calculation. In the case of the PDLD model where each group is taken as a Langevin dipole, we only have to consider the product of each dipole and the corresponding field term of the fourth-order expansion.[17] The field of eq. (1) is now rewritten as

$$\xi_i = \xi_i^o + (\xi_i^s + \xi_i^l) - \xi_i^c \tag{9}$$

where ξ^s is the short-range field due to the Langevin dipoles within the cutoff radius and ξ^l is the long-range field due to the rest of the Langevin dipoles of the system so that $\xi^s + \xi^l = \xi^L$. While ξ^s is evaluated at every iteration, ξ^l is updated only once in every 10 iterations (in general, 25 iterations are needed for convergence). This rather simple modification not only gives the corresponding results obtained without any cutoff but also increases significantly the performance of the POLARIS program for large systems with many (>1000) solvent molecules.

Averaging the PDLD Energy over Protein Configurations

Early studies[3,9,10] evaluated the PDLD energies at the average solute configuration. This implies that the average of the electrostatic energy over many solute configurations can be approximated by the electrostatic energy at the average structure (taken usually from the X-ray structure). However, the recent increase in computer power allows one to consider an actual average over configurations generated by MD simulations at low temperature.[18] This provides a more rigorous expression for the free energy, ΔG_{elec}, associated with changing the charge configuration of the solute from one charge state (state A) to another charge state (state B), taking into account consistently the reorganization of the system during the change in its charge state (see ref. 18). With the current averaging over the configurations of state B, we can express ΔG_{elec} as

$$\begin{aligned}\Delta G_{elec} &= \Delta G(Q_A, \mathbf{r}_A \longrightarrow Q_B, \mathbf{r}_B) \\ &= \langle \Delta V_{QQ} + \Delta V_{Q\mu} + \Delta V_{Q\alpha} + \Delta G_{Qw} \rangle_B \\ &\quad + \langle \Delta\Delta G_{hyd} \rangle_B + \lambda + \Delta G_{is} \\ &= \langle U_{elec}^{eff} \rangle_B + \lambda + \Delta G_{is} \end{aligned} \tag{10}$$

where Q and $\mathbf{r}$ designate, respectively, the charge distribution and the coordinates of the indicated state. $\langle\ \rangle_B$ designates an average of the indicated function of the equilibrium distribution of the Bth state. The averaging procedure is done automatically by sending the molecular configurations generated each half picosecond along the trajectory to the PDLD part of POLARIS. This approach provides much more accurate free energies (e.g., binding energies) than those obtained from single protein configurations. The terms $\Delta\Delta G_{hyd}$, λ, and ΔG_{is} will be considered below.

Field-Dependent Hydrophobic Correction

The term $\Delta\Delta G_{hyd}$ of eq. (10) is the *change* in the hydrophobic free energy ΔG_{hyd} due to the charging process, which can be estimated by the recently developed approach[16] that involves a crucial modification of our early approach.[19] The modified treatment estimates the effect of the field from the solute on the hydrophobic energy. The field corrected hydrophobic energy is given by

$$\Delta G_{hyd} = -T\Delta S_{hyd} \tag{11}$$

$$\Delta S_{hyd} = A + B \sum_{i \in s} CF(\xi_i^o)$$

$$F(\xi_i^o) = \begin{cases} 1 & \xi_i^o < \xi_{phobic} \\ \left(1 - \dfrac{\xi_i^o - \xi_{phobic}}{\xi_{philic} - \xi_{phobic}}\right) & \xi_{phobic} < \xi_i^o < \xi_{philic} \\ 0 & \xi_i^o > \xi_{philic} \end{cases}$$

where i runs over the sites of the Langevin dipoles on the surface (s) of the solute. F is a function of the electric field, ξ_i^o, on the ith Langevin dipole at the solvation surface. F is 1 when the field is small ($<\xi_{phobic}$) and 0 when the field is large ($>\xi_{philic}$). A, B, C, ξ_{philic}, and ξ_{phobic} are scaling constants obtained from studies of the free energies of transfer of different molecules from hydrocarbons to water.[16] This approach was found to give excellent results in recent studies of solvation entropies of polar and nonpolar molecules.[16]

Reorganization Energy Correction as an Estimate of Sterically Induced Electrostatic Effects

The term λ of eq. (10) represents the "reorganization energy" associated with the change of the protein

structure between $\mathbf{r}_A$ and $\mathbf{r}_B$,[20] which can be estimated by the approach of ref. 18. This approach approximates λ by

$$\lambda \simeq \frac{1}{2}(\langle U_{elec}^{eff}\rangle_A - \langle U_{elec}^{eff}\rangle_B) \tag{12}$$

where $\langle U_{elec}^{eff}\rangle_A$ and $\langle U_{elec}^{eff}\rangle_B$ are evaluated with the two minimized protein structures, $\mathbf{r}_A$ and $\mathbf{r}_B$, that correspond to the configurations obtained with the uncharged and the charged ligand, respectively. It is important to realize that the reorganization energy reflects the *steric effect* of the protein. Apparently, one can consider the steric forces as the restoring forces for the electrostatic interactions in systems that follow the linear response approximation (see Appendix 5 of ref. 3 for a related discussion).

Calculations of Ionic Strength Effects

The last term ΔG_{is} of eq. (10) represents the effect of ionic strength associated with the change of the solute charges from Q_A to Q_B. In general, the effect of ionic strength on rate constants and binding is small in proteins (i.e., <1 kcal mol^{-1} as compared to overall electrostatic contributions in enzyme catalysis and other biologic processes, which can be around 10 kcal mol^{-1}). Further, because the effect of the ionic strength involves ions in water it can be estimated quite reliably using macroscopic models including discretized continuum treatments of the Poisson–Boltzmann equation[6] and Debye–Huckel type treatments.[21] Thus, we take a two-step strategy in our calculations. The first step involves the regular microscopic PDLD calculation of the electrostatic energy of the solute in the absence of any ionic atmosphere. The second step involves a macroscopic estimate of the electrostatic interaction between the solute and protein groups in regions 1 and 2 and the given ionic atmosphere. The best way to understand this approach is to consider the interaction between an ionizable group of the protein and an ion in water. This interaction can be evaluated by first calculating the self-energy of the given group in water at infinite separation from the given ion and then calculating the *change* in energy ($\Delta\Delta G_{sol} + \Delta V_{QQ}$) upon bringing the ion from infinity to its given location. However, the energy associated with the second step can also be estimated on a macroscopic level using Coulomb's law and a large dielectric constant. The same considerations can be applied to many (rather than just one) ions in water.

Our specific macroscopic treatment for the above second step is a variant of the grid approach introduced by Pack and coworkers (e.g. ref. 22). This treatment represents the ionic atmosphere (within a specific radius from the center of our system) by a grid of residual charges. The grid spacing is taken here as 3 Å and the net charge density within each cubic element centered at the ith grid point is determined by considering the corresponding Boltzmann distribution and writing

$$\rho_i = \rho_i^+ - \rho_i^- = N_o\left[\frac{\exp(-\beta\Phi_i)}{\Sigma_k \exp(-\beta\Phi_k)} - \frac{\exp(\beta\Phi_i)}{\Sigma_k \exp(\beta\Phi_k)}\right] \tag{13}$$

where ρ_i^+ and ρ_i^- are, respectively, the charge densities of the cations and anions in the ith element, β is $1/k_BT$, N_o is the number of cations or anions per volume element, and Φ_i is the potential at the ith grid point due to the charges of the system and the electrolytes, given in our macroscopic model by

$$\Phi_i = \sum_j \frac{q_j^P}{\epsilon_{ij} r_{ij}} + \sum_{k\neq i} \frac{q_k^G}{\epsilon_{ik} r_{ik}} \tag{14}$$

Here, q_j^P is the residual charge of the jth protein atom and q_k^G is the point charge at the kth grid point (representing the excess net charge of the kth volume element). The ϵ_{ij} of eq. (14) can be approximated by a large number between 40 and 80 or by the function (see ref. 7)

$$\epsilon(r_{ij}) = 1 + 60[1 - \exp(-0.1 r_{ij})] \tag{15}$$

(a similar function has been adopted recently in the study of ref. 23). The residual charge at the kth grid point is now given by

$$q_k^G = \tau_k \rho_k \tag{16}$$

where τ_k is the volume of the kth element. The final set of the grid charges is obtained by solving eqs. (13) and (14) iteratively. With these q^Gs, we can now evaluate the effect of the ionic strength, ΔG_{is}, by

$$\Delta G_{is} = 332 \sum_{i,j} \frac{q_i^P q_j^G}{\epsilon_{ij} r_{ij}} + 332 \sum_{j,k>j} \frac{q_j^G q_k^G}{\epsilon_{jk} r_{jk}} \tag{17}$$

where q_i^P is the residual charge of the ith protein atom and q^G is the point charge at the corresponding grid point.

Calculations of Titration Curves

Although the effects of ionizable surface groups is in general small, it is important to estimate such effects. This is in particular true when one is interested in the dependence of the properties of the given protein upon the external pH. Our treatment of this problem is based upon a two-step procedure. In the first step, we use the PDLD or the PDLD/S method to evaluate the intrinsic pK_as of all ionizable groups. This is done by evaluating the self-energy of each group where all other groups are at their neutral states. In the second step, we evaluate the average ionization state of each group considering the interaction between the different groups on a macroscopic level.

A formally rigorous approach for the second step is provided by the approach of refs. 24 and 25, where the energetics of all the possible charge configurations is expressed by

$$W^{(m)} = -\sum_i 2.3RTq_i^{(m)}[pK_{intl}(A_i) - pH] + \frac{1}{2}\sum_{i\neq j} q_i^{(m)}q_j^{(m)}/(R_{ij}^{(m)}\epsilon_{ij}^{(m)}) = \sum_i U_i^o q_i^{(m)} + \frac{1}{2}\sum_{i\neq j} U_{ij}q_i^{(m)}q_j^{(m)} \quad (18)$$

where the $q_i^{(m)}$ can be 0 or -1 for acids and 0 or 1 for bases. The ϵ_{ij} of eq. (18) can be approximated by the ϵ of eq. (15) as established in many studies of charge–charge interactions in proteins (e.g., refs. 3 and 7).

Now, we can express the total free energy and the average charges by

$$\Delta G = \sum_m W^{(m)}P^{(m)} \quad (19)$$

$$P^{(m)} = e^{-W^{(m)}\beta} \Big/ \sum_m e^{-W^{(m)}\beta}$$

$$\langle q_i\rangle = \sum_m q_i^{(m)}P^{(m)}$$

The sum over all configurations might involve extensive computer time and be replaced by important sampling approaches (see ref. 26 for a related Monte Carlo study).

While the above option is appealing, we adopted in the current version of POLARIS a simpler and less accurate approach, considering the q's as fractional charges and writing

$$\Delta G \simeq \sum_i U_i^0\langle q_i\rangle + \frac{1}{2}\sum_{i\neq j} U_{ij}\langle q_i\rangle\langle q_j\rangle \quad (20)$$

This macroscopic approximation (consistent in some respects with the Poisson–Boltzmann and related approaches) estimates the reversible work of bringing the fractional charges from infinity to their given protein site. With this formulation, we estimate the charge $\langle q_i\rangle$ by freezing all other fractional charges at their current value and using the two states probability:

$$\langle q_i\rangle = Q_i^o e^{-\Delta G(q_i=0\to Q_i^o)\beta}/(e^{-\Delta G(q_i=0\to Q_i^o)\beta} + e^{+\Delta G(q_i=0\to Q_i^o)\beta}) \quad (21)$$

where Q_i^o is the charge of the ith group in its ionized form. Equation (21) is solved iteratively in a convenient way allowing the charges to change by 0.3 units at each iteration.

PDLD Parameters

The van der Waals parameters of the PDLD model, being used in the present version of the POLARIS program, are given in Table I. These parameters were obtained by the same calibration procedure introduced in the early PDLD studies (e.g., ref. 19). The charges used in the PDLD calculations are the same as those used in the all-atom calculations and are given in Tables II–IV. It is important to note that the philosophy of calibrating van der Waals radii using observed solvation energies of representative molecules is now common to many FEP approaches. In fact, the most reliable treatments are based upon this approach rather than upon *ab initio* determination of the relevant parameters. It is crucial, of course, to keep the van der Waals parameters *unchanged* when applied to different ligand or protein molecules.

Table I. van der Waals parameters of the PDLD model.

Atom	σ
H	1.50
C(*sp*2)	2.00
Ca	2.01
B(*sp*)	1.35
Br	2.50
Zn	1.42
Sb	1.76
Tc	2.00
O	2.00
P	3.20
Mg	1.50
B(*sp*2)	1.23
I	3.00
Fe	1.10
Se	1.60
Te	3.10
N	2.40
S	3.10
Al	0.90
F	2.00
Ga	1.10
In	1.81
Si	2.80
Ti	1.42
C(*sp*3)	2.40
Na	2.10
As	1.58
Cl	2.50
Ge	1.60
Ru	1.67
Sn	1.80

The van der Waals energy (in kcal mol^{-1}) for the PDLD model is given by $V_{vdw} = (\epsilon_i\epsilon_j)^{1/2}[((\sigma_i + \sigma_j)/2)^{12}r_{ij}^{-12} - 2((\sigma_i + \sigma_j)/2)^6 r_{ij}^{-6}]$, where i and j denote the atoms in the pairwise interaction and r_{ij} is the distance between the ith and jth atoms. ϵ_i is taken as 0.3 kcal mol^{-1} for all atom types, σ (in Å) corresponds to the cutoff radius for the distance between the given atom type and the closest Langevin dipole, while *sp*, *sp*2, and *sp*3 represent the types of orbital hybridization for the given atom type.

Table II. Partial atomic charges for some common amino acids.

Main chain and Gly											
	N	H	C_α	H_α	C'	O					
	−0.400	0.400	−0.097	0.097	0.550	−0.550					
Side chains of polar amino acids											
Asp^-	C_β	H_β	C_γ	O_{δ_1}	O_{δ_2}						
	−0.194	0.097	0.360	−0.680	0.680						
Ser^-	C_α	H_α	C_β	H_β	O_γ	H_γ					
	−0.050	0.050	−0.100	0.050	−1.000	0.000					
His^+	C_β	H_β	C_γ	N_{δ_1}	H_{δ_1}	C_{δ_2}	H_{δ_2}	C_{ϵ_1}	H_{ϵ_1}	N_{ϵ_2}	
	−0.194	0.097	0.160	−0.160	0.190	0.540	0.070	0.100	0.070	0.030	
Ser	C_α	H_α	C_β	H_β	O_γ	H_γ					
	−0.050	0.050	−0.100	0.050	−0.450	0.450					
His	C_β	H_β	C_γ	N_{δ_1}	H_{δ_1}	C_{δ_2}	H_{δ_2}	C_{ϵ_1}	H_{ϵ_1}	N_{ϵ_2}	
	−0.194	0.097	0.035	0.013	0.187	0.058	0.070	0.110	0.077	−0.550	
Asp	C_β	H_β	C_γ	O_{δ_1}	O_{δ_2}						
	−0.194	0.097	0.450	−0.450	0.000						
Arg	C_β	H_β	C_γ	H_γ	C_δ	H_δ	N_ϵ	H_ϵ	C_ζ	N_η	H_η
	−0.194	0.097	−0.194	0.097	−0.194	0.097	−0.253	0.253	0.000	−0.834	0.417
Asn	C_β	H_β	C_γ	O_{δ_1}	N_{δ_2}	H_δ					
	−0.194	0.097	0.500	−0.500	−0.834	0.417					
Lys	C_β	H_β	C_γ	H_γ	C_δ	H_δ	C_ϵ	H_ϵ	N_ζ	H_ζ	
	−0.194	0.097	−0.194	0.097	−0.194	0.097	−0.194	0.097	−1.080	0.360	
Gln	C_β	H_β	C_γ	H_γ	C_δ	O_{ϵ_1}	N_{ϵ_2}	H_ϵ			
	−0.194	0.097	−0.194	0.097	0.500	−0.500	−0.834	0.417			
Thr	C_α	H_α	C_β	H_β	O_{γ_1}	H_{γ_1}	C_{γ_2}	H_{γ_2}			
	−0.050	0.050	0.050	0.050	−0.600	0.450	−0.241	0.097			
Glu	C_β	H_β	C_γ	H_γ	C_δ	O_{ϵ_1}	O_{ϵ_2}				
	−0.194	0.097	−0.194	0.097	0.400	−0.400	0.000				
Side chains of nonpolar amino acids											
Ala	C_β	H_β									
	−0.291	0.097									
Val	C_β	H_β	C_{γ_1}	H_{γ_1}	C_{γ_2}	H_{γ_2}					
	−0.097	0.097	−0.291	0.097	−0.291	0.097					
Leu	C_β	H_β	C_γ	H_γ	C_{δ_1}	H_{δ_1}	C_{δ_2}	H_{δ_2}			
	−0.194	0.097	−0.097	0.097	−0.291	0.097	−0.291	0.097			
Ile	C_β	H_β	C_{γ_1}	H_{γ_1}	C_{γ_2}	H_{γ_2}	C_δ	H_δ			
	−0.097	0.097	−0.194	0.097	−0.291	0.097	−0.291	0.097			
Pro	C_β	H_β	C_γ	H_γ	C_δ	H_δ					
	−0.194	0.097	−0.194	0.097	0.194	0.097					
Phe[a]	C_β	C_γ	C_{δ_1}	C_{ϵ_1}	C_ζ	C_{ϵ_2}	C_{δ_2}				
	−0.194	0.000	−0.097	−0.097	−0.097	−0.097	−0.097				
Tyr[a]	C_β	C_γ	C_{δ_1}	C_{ϵ_1}	C_ζ	O_η	H_η	C_{ϵ_2}	C_{δ_2}		
	−0.194	0.000	−0.097	−0.097	0.150	−0.600	0.450	−0.097	−0.097		
Trp[a]	C_β	C_γ	C_{δ_1}	N_{ϵ_1}	H_{ϵ_1}	C_{ϵ_2}	C_{ζ_2}	C_{η_2}	C_{ζ_3}	C_{ϵ_3}	C_{δ_2}
	−0.194	0.000	−0.097	−0.450	0.450	0.000	−0.097	−0.097	−0.097	−0.097	0.000
Cys	C_α	H_α	C_β	H_β	S_γ	H_γ					
	−0.050	0.050	0.050	0.050	−0.600	0.450					
Met	C_β	H_β	C_γ	H_γ	S_δ	C_ϵ	H_ϵ				
	−0.194	0.097	0.056	0.097	−0.500	−0.041	0.097				

The charges are given in atomic units.
[a]The partial atomic charges for the rest of all the hydrogens are 0.097.

Table III. Partial atomic charges for some protonated amines.

Amine	N	H	C_α	H_α	C_β	H_β
NH_4^+	−0.080	0.270				
$CH_3NH_3^+$	0.000	0.269	0.022	0.057		
$C_2H_5NH_3^+$	−0.046	0.272	0.080	0.062	−0.100	0.042
$(CH_3)_3NH^+$	0.294	0.220	−0.009	0.057		

The table gives the atomic charges for the protonated amines examined in this work. The charge is given in atomic units.

Table IV. Partial atomic charges for some other organic molecules.

CH_4	C	H								
	−0.160	0.040								
C_5H_{12}	C_α	H_α	C_β	H_β	C_γ	H_γ	C_δ	H_δ	C_ϵ	H_ϵ
	−0.120	0.037	−0.070	0.038	−0.070	0.038	−0.070	0.038	−0.120	0.037
C_6H_6	C	H								
	−0.050	0.050								
CH_3OH	H	O	C_α	H_α						
	0.450	−0.450	−0.114	0.038						
CH_3CH_2OH	H	O	C_α	H_α	C_β	H_β				
	0.450	−0.450	−0.066	0.033	−0.114	0.038				
C_6H_5OH	H	O	C_α	C_β	H_β	C_γ	H_γ	C_ϵ	H_ϵ	
	0.450	−0.450	0.030	−0.064	0.050	−0.057	0.053	−0.048	0.053	
C_5H_5N	N	C_α	H_α	C_β	H_β	C_γ	H_γ			
	−0.511	0.160	0.061	−0.044	0.054	−0.006	0.053			
$C_5H_5NH^+$	H	N	C_α	H_α	C_β	H_β	C_γ	H_γ		
	0.210	−0.120	0.240	0.060	0.010	0.060	0.110	0.060		
$C_6H_5NH_2$	H	N	C_α	C_β	H_β	C_γ	H_γ	C_ϵ	H_ϵ	
	0.417	−0.834	0.000	−0.050	0.050	−0.050	0.050	−0.050	0.050	

Atomic charges for the other organic molecules examined in this work. The charges are given in atomic units.

Implementation of the Surface-Constrained All-Atom Model and the FEP Method in ENZYMIX

Although the simplified solvent representation of the PDLD model seems to give reasonable results,[4] it is important to examine the performance of all-atom solvent models. Such treatments, which are in principle more rigorous, involve major convergence problems and require the inclusion of a large number of solvent molecules. One way of obtaining reasonable results with a limited number of solvent molecules is to use spherical boundary conditions with special surface constraints. Such constraints (e.g., refs. 3, 20, and 27–30) should force the finite system to behave as the corresponding region in an infinite system. The present version is called the Surface-Constrained All-Atom Solvent (SCAAS) model.[3,27,28] This approach emphasizes the electrostatic constraint, forcing the polarization and compression of the finite simulation system, in response to the field due to internal charges, to approximate the polarization expected from an infinite system. Other approaches (e.g., those described in ref. 29) emphasize the correct heat transfer and density at the boundaries of the system and do not guarantee that the electrostatic response of the finite system will follow that of the complete system (see the discussion in ref. 28). It is also important to recognize that the frequently used periodic boundary conditions do not have the proper symmetry for treatment of ions (see refs. 28 and 29). The present version of the SCAAS approach, implemented in the program ENZYMIX,[15] focuses on obtaining a reliable treatment of long-range forces. This is accomplished by dividing the protein into five regions as depicted in Figure 3. Region 1 contains the groups whose electrostatic energy is of interest. Region 2 contains the unconstrained protein electroneutral and charged groups selected within a radius R_2 from the designated center. Region 3a contains the unconstrained water molecules generated within a radius R_3 and region 3b contains the surface water molecules (within R_3) that are subjected to a set of angular and radial constraints designed to keep the average polarization and density of region 3a as close as possible to those expected from the corresponding infinite system (see ref. 28 for more details of our surface constraint approach). Regions 2 and 3 are surrounded by region 4a, which includes the protein groups between R_2 and R_{4a}, and by a spherical grid of Langevin dipoles (region 4b) of a radius R_{4b}. The protein groups in region 4a are constrained to the corresponding X-ray positions by quadratic constraints. Region 4 (4a and 4b) is then surrounded by a bulk region, region 5, whose electrostatic effects are simulated by considering it as a continuum with the dielectric constant of bulk water. While the electrostatic effect of the protein part of region 5 (region 5a) is represented by a dielectric continuum, the steric and bonding effects of this region are treated explicitly (this helps maintain the protein in its correct folded structure). The introduction of regions 4 and 5 is a new feature of our all-atom model. Inclusion of the Langevin grid allows us to generate system of explicit water dipoles far beyond R_3 and obtain in a relatively inexpensive way more realistic long-range solvent effect than those obtained by a simple continuum treatment [see eq. (8)]. The spherical truncation of the protein allows us to obtain the continuum contribution in a stable way and avoid artifacts associated with incorrect treatments of the long-range effects.

It should be noted that both the energy and the forces from the Langevin grid are included in the simulation (in treating the bulk contribution, we consider only the energy and neglect the corresponding forces). The incorporation of the Langevin grid in the all-atom model of the FEP approach is similar

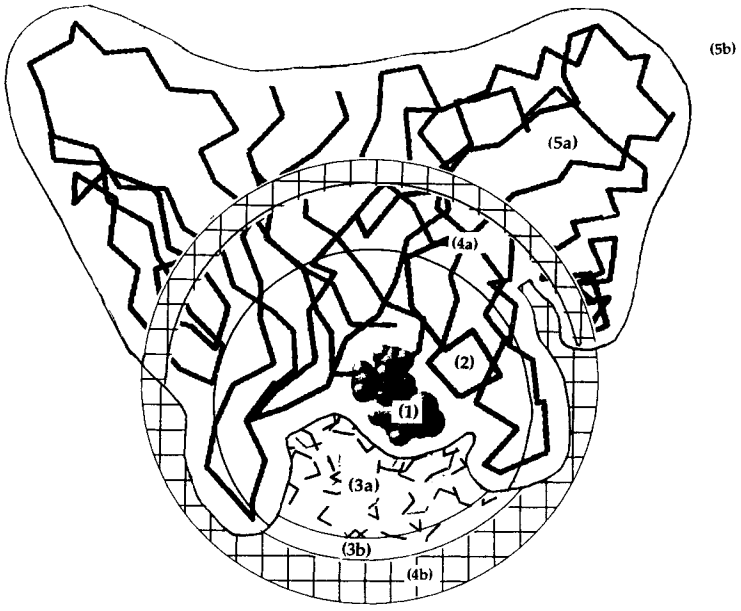


Figure 3. Five regions of the protein system in the surface-constrained all-atom model used in the FEP calculations. Region 1 contains the charged groups of interest. Region 2 contains the unconstrained protein groups within a radius R_2 from the center. Region 3a contains the unconstrained water molecules and region 3b contains the surface water molecules (which are subjected to polarization constraints). Region 4a contains the harmonically constrained protein atoms between the radii R_2 and R_{4a}, and region 4b contains the grid of Langevin dipoles that complete regions 1, 2, 3, and 4a to a sphere of a radius R_{4b}. Region 5 (which includes region 5a of the protein) is modeled by considering the entire region as a dielectric continuum.

to that of the regular PDLD treatment. That is, a cubic *rigid* grid with a spacing of 4 Å is built around regions 1, 2, 3, and 4a of the protein system (the use of a 4-Å grid spacing and the corresponding use of a larger dipoles give results similar to those obtained with a 3-Å grid spacing but with fewer dipoles and less computer time). The grid is then truncated to a sphere of a radius R_{4b} from the center. Grid points closer than 3 Å to any atoms in those regions are deleted.

The local field on the ith the Langevin dipole is expressed in a similar way to that used in eq. (1), i.e.,

$$\boldsymbol{\xi}_i = (\boldsymbol{\xi}_i^o + \boldsymbol{\xi}_i') + \boldsymbol{\xi}_i^L \tag{22}$$

where $\boldsymbol{\xi}_i^o$ is the contribution from the permanent charges in regions 1, 2, and 3 (which include both protein and water atoms), $\boldsymbol{\xi}_i'$ is the contribution from the protein atoms in region 4a (evaluated only once in a simulation because these atoms are effectively fixed by constraints), and $\boldsymbol{\xi}_i^L$ is the contribution from *all* other Langevin dipoles (i.e., nearest neighbors are also included in this case). Both $\boldsymbol{\xi}_i^o$ and $\boldsymbol{\xi}_i'$ are scaled by the same function $f_d(r_{ij})$ of eq. (7). For the contribution from protein groups, r_{ij} of eq. (7) is taken as the distance between the jth protein atom and the ith Langevin dipole, while for the contribution from water molecules it is taken as the distance between the jth water oxygen and the ith Langevin dipole. The parameter r_d is also taken as 3 Å.

The average polarization of the ith Langevin dipole, $\boldsymbol{\mu}_i^L$, by its local field is given by eq. (2). The electrostatic interaction between the Langevin dipoles and the charges in regions 1, 2, and 3 is evaluated by

$$\Delta G = -332 \sum_i \boldsymbol{\mu}_i^L \boldsymbol{\xi}_i^o \tag{23}$$

The van der Waals interaction between the Langevin dipoles and the relevant atoms is determined by a simple repulsive potential that varies as R^{-12}. Note that the factor ½ of eq. (8) is not included in eq. (23). The reason for this is that we include the Langevin contribution as an effective potential in the FEP calculations and the penalty for polarizing the dipoles is evaluated explicitly in such calculations. The force on the ith protein atom, $\mathbf{f}_i$, from the jth Langevin dipole is evaluated by minus the gradient of the corresponding potential energy, i.e.,

$$\mathbf{f}_i = -\mathbf{f}_j = -332\, \boldsymbol{\mu}_j^L (\nabla_i \cdot \boldsymbol{\xi}_{ij}^o) \tag{24}$$

where $\mathbf{f}_j$ is the corresponding force on the jth Langevin dipole (which is not needed because the grid is rigid), $\boldsymbol{\mu}_j^L$ is the jth Langevin dipole from the last iteration, $\boldsymbol{\xi}_{ij}^o$ is the field on the jth Langevin dipole from the ith protein atom, and ∇_i is the gradient operator with respect to the coordinates of the ith atom.

Because the Langevin dipoles of our model are usually far away from the region of interest (i.e.,

region 1), their collective long-range effect can be considered as a slowly varying function. Thus, we update the field ξ^0 of eq. (22) and the field-gradient only once in every 10 time steps (i.e., 20 fs). In other words, the energy and forces from the Langevin dipoles are kept constant during this time period. However, the field ξ^L of eq. (22) is updated at every step and is used to update μ^L [by eq. (2)], thus allowing for self-consistency in the evaluation of the Langevin dipoles. This rather simple modification is found to substantially increase the speed of the calculations without affecting the accuracy and precision of the results.

In addition to the forces associated with the effect of the Langevin dipoles on regions 1, 2, and 3, the regular forces in these regions are considered as well. However, the forces due to the induced dipoles of the protein atoms and the water molecules are not considered explicitly and their effects are approximated by reducing the magnitudes of the coulombic forces by a factor of 0.5 (see Fig. 13 in ref. 3 for the rationale behind this approximation). Similar results are obtained by the noniterative induced dipole treatment,[9,27] which is also integrated conveniently in ENZYMIX. Despite the introduction of the new boundary conditions and the external Langevin grid, it is still possible to get incorrect electrostatic energies by our model if one uses regular cutoff distances (e.g., 8 Å) for the electrostatic interactions between the atoms in regions 1, 2, and 3 (see ref. 31 for a demonstration of this point). Thus, it is essential to use a large cutoff or no cutoff for the above interactions and this might require a long simulation time. Here, we find the introduction of the LRF method[17] in the ENZYMIX program extremely useful, that is, although the LRF treatment of all-atom systems requires a fourth-order expansion (rather than the simple treatment needed for the LRF treatment of the PDLD model) it leads to results that are similar to those obtained without any cutoff but in much shorter time (see ref. 17).

The charges and nonbonded parameters of the model were calibrated by requiring them to reproduce observed solvation free energies. The van der Waals parameters of the all-atom model are given in Table V. The atomic charges of the all-atom model for some common amino acids and some other organic molecules are given in Tables II–IV.

With a reasonable all-atom treatment, we can turn to evaluate the free energies of our system. A particularly powerful strategy is provided by the FEP approach. In this approach, one determines the free energy difference between two states, A and B, by using an effective potential that drives the system from one state to another in small steps, so that the free energy change in each step can be evaluated by a thermodynamic perturbation approach.[32,33] Studies of charge transfer reactions and other solvation processes in solutions and proteins[27,34–39] have suggested that this can be done effectively by a potential of the form

Table V. Some common van der Waals parameters of the all-atom model.

Atom	A	B
H	22.0	1.0
H (in H_2O)	0.4	0.6
O	793.3	25.0
O (in H_2O)	861.4	26.0
O^-	2066.0	42.0
N	1425.0	42.0
C	1956.0	32.0
P	2454.0	47.0
S	1831.0	40.5
F^-	800.0	2.5
Cl^-	7000.0	2.5
Br^-	8000.0	2.5
I^-	16,000.0	2.5
Li^+	17.3	1.2
Na^+	150.0	5.0
K^+	730.0	14.0
Rb^+	1260.0	20.0
Cs^+	2250.0	30.0
Ca^{2+}	345.0	15.0

The van der Waals energy (in kcal mol^{-1}) for the all-atom model is given by $V_{vdw} = A_iA_jr_{ij}^{-12} - B_iB_jr_{ij}^{-6}$, where i and j denote the atoms in the pairwise interaction and r_{ij} is the distance between the ith and jth atoms. A_i and B_i are given in units of $Å^6$ $kcal^{1/2}$ $mol^{-1/2}$ and $Å^3$ $kcal^{1/2}$ $mol^{-1/2}$, respectively.

$$\varepsilon_m = (1 - \theta_m)\varepsilon_A + \theta_m\varepsilon_B \quad (25)$$

where ε_A and ε_B are the potential surfaces of state A and state B, respectively. θ_m is a mapping parameter used to change ε_m from ε_A to ε_B. For small steps in θ, the free energy change associated with changing θ_m to $\theta_{m'}$ can be treated as a perturbation on the ensemble belonging to θ_m, i.e.,

$$\delta G(\theta_m \longrightarrow \theta_{m'}) = -RT \ln\langle\exp\{-(\varepsilon_{m'} - \varepsilon_m)/RT\}\rangle_m \quad (26)$$

The total free energy difference between the states A and B is given by the sum of the contributions from each θ step:

$$\Delta G(\theta_0 \longrightarrow \theta_n) = \sum_{m=0}^{n-1} \delta G(\theta_m \longrightarrow \theta_{m'}) \quad (27)$$

The use of this formally correct equation for the evaluation of absolute solvation energies of charges in proteins is a major challenge in terms of convergence and proper boundary conditions. This point, which is only now being generally appreciated, will be addressed in the section on examining model performance.

Scaled Microscopic Models Can Extend the Precision of Electrostatic Calculations: The PDLD/S Method

The main problem with the above microscopic approaches is related to the fact that such models involve the sum of large contributions (e.g., $V_{Q\mu}$ and

V_{QQ}) that even with a small relative error lead to a significant absolute error. On the other hand, the macroscopic picture is associated with much smaller energy contributions because the relevant interactions are divided by a large dielectric constant. This difference between the microscopic and macroscopic approaches is illustrated in Figure 4 for an ion pair in water. As seen from the figure, the microscopic approach gives small energy changes for moving the charges from infinity to R through an almost perfect compensation of large number (V_{QQ} and ΔG_{sol}). In the macroscopic picture, the above compensation is basically *assumed* by using a large dielectric constant. In other words, the macroscopic approach integrates the weak effective force (which is the field scaled by the dielectric constant) for moving the charges in solution, while the microscopic model evaluates directly the energy of moving the charges to their given site from *vacuum*. The problem with the macroscopic approach appears, however, when one deals with moving charges between different phases (e.g., from water to a protein). Now, in addition to integrating the forces, one must also take into account the *change in self-energy* that accompanies the movement of the charges from one phase to another. Unfortunately, this aspect is needed for the accuracy rather than the precision of the evaluation of the energies of charges in proteins.

Considering the higher precision of the macroscopic approaches, one may look for related ways of scaling large energy contributions of the microscopic models, obtaining similar precision, and yet retaining the clear energy-based concepts of these models. Such a "scaled microscopic" model, referred to here as the PDLD/S model, has been introduced[4,40] and applied recently[31] and its examination is one of the primary objectives of the present work. Probably

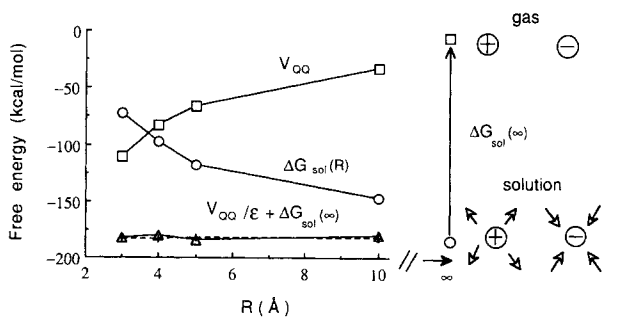


Figure 4. Relationship between microscopic and macroscopic descriptions of the energetics of an ion pair. The figure gives the energy of an ion pair as a function of the interatomic distance, evaluated macroscopically by $V_{QQ}/\epsilon + \Delta G_{sol}(\infty)$ (-+-), and the corresponding microscopic energy, evaluated by $V_{QQ} + \Delta G_{sol}$ (-Δ-). The plot demonstrates the remarkable compensation of the changes in the charge–charge interaction V_{QQ} (-□-) by changes in the solvation energy ΔG_{sol} (-○-). $\Delta G_{sol}(\infty)$ denotes the sum of the solvation energies of the ions at infinite separation.

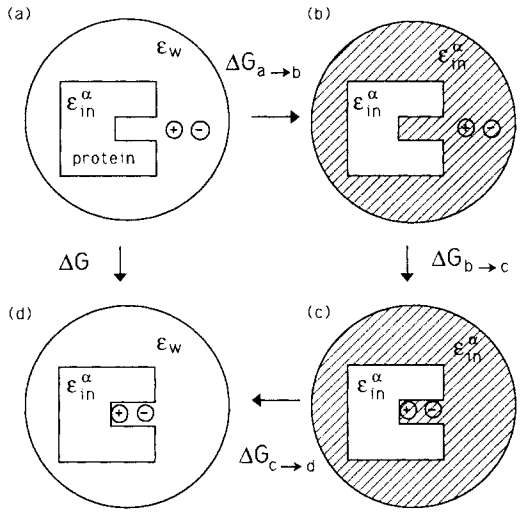


Figure 5. Thermodynamic cycle used in the PDLD/S method for calculating the free energy of charges in a protein. See the text for details.

the best way to understand the PDLD/S scaling procedure is to follow the thermodynamic cycle depicted in Figure 5. The first step of the cycle starts with the relevant charged groups of the protein immersed in water at infinite separation from each other, while the sites of these groups in the protein are occupied by neutral groups of the same size. The electrostatic energy of the system at this state is given by

$$\Delta G_a = \sum_i \Delta G^i_{sol,w} + \Delta G^p_{sol,w}(\mathbf{Q} = 0) \quad (28)$$

where $\mathbf{Q}$ designates the vector of the protein charges, $\Delta G^i_{sol,w}$ is the PDLD solvation free energy of the ith charged group in water, and $\Delta G^p_{sol,w}(\mathbf{Q} = 0)$ is the PDLD solvation energy of the protein with all the Qs in their uncharged forms (a more accurate treatment should involve empty cavities instead of the uncharged groups[31]). The second step involves changing the dielectric constant of the entire water region from ϵ_w to a dielectric constant, ϵ^{α}_{in}, whose meaning will be discussed below. The free energy change associated with this step can be approximated by

$$\Delta G_{a\to b} \simeq -\left[\sum_i \Delta G^i_{sol,w} + \Delta G^p_{sol,w}(Q = 0)\right] \times \left(\frac{1}{\epsilon^{\alpha}_{in}} - \frac{1}{\epsilon_w}\right) \quad (29)$$

This expression reflects the fact that the change in the macroscopic solvation energy upon moving a molecule from ϵ_w to ϵ^{α}_{in} is given by scaling the corresponding energy in ϵ_w by $1/\epsilon^{\alpha}_{in}$ for $\epsilon_w >> \epsilon^{\alpha}_{in}$. Note that in the macroscopic treatment the Born's energy for moving an ion from ϵ_w to ϵ^{α}_{in} is scaled by $(1/\epsilon^{\alpha}_{in} - 1/\epsilon_w)/(1 - 1/\epsilon_w)$, which can be approximated by $1/\epsilon^{\alpha}_{in}$ for $\epsilon_w >> \epsilon^{\alpha}_{in}$, and a similar approx-

imation is applicable to the dipole contribution. The $\Delta G^{i}_{sol,w}$ terms are evaluated microscopically in our treatment rather than by the Born's formula with arbitrary cavity radii. Yet, we retain the $1/\epsilon^{\alpha}_{in}$ macroscopic scaling.

In the next step, we move the Qs from infinity to their actual protein sites. The free energy change of this step is given by

$$\Delta G_{b\to c} = (V_{QQ} + V_{Q\mu})\frac{1}{\epsilon^{\alpha}_{in}} \tag{30}$$

where V_{QQ} is the vacuum interaction between the charges and $V_{Q\mu}$ is the vacuum interaction between the charges and the protein polar groups. Here, we exploit the fact that the energy of bringing charges from infinity to a given distance at a uniform medium is given by scaling the vacuum interaction energy by the dielectric constant of that medium.

The last step that completes our cycle is the change of the dielectric constant of the medium surrounding the protein from ϵ^{α}_{in} back to ϵ_w. The free energy change associated with this stage can be approximated by

$$\Delta G_{c\to d} \simeq \Delta G^{p}_{sol,w}(\mathbf{Q} = \mathbf{Q}_o)\left(\frac{1}{\epsilon^{\alpha}_{in}} - \frac{1}{\epsilon_w}\right) \tag{31}$$

Now, the total free energy of bringing the charged groups from water to their protein sites is given by

$$\begin{aligned}\Delta G &= \Delta G_{a\to d}\\ &= \Delta G_{a\to b} + \Delta G_{b\to c} + \Delta G_{c\to d}\\ &= \left[\left(\sum_i - \Delta G^{i}_{sol,\infty}\right) + \Delta G_{Qw}\right]\left(\frac{1}{\epsilon^{\alpha}_{in}} - \frac{1}{\epsilon_w}\right)\\ &\quad + (V_{QQ} + V_{Q\mu})\frac{1}{\epsilon^{\alpha}_{in}}\end{aligned} \tag{32}$$

where $\Delta G_{Qw} = \Delta G^{p}_{sol,w}(\mathbf{Q} = \mathbf{Q}_o) - \Delta G^{p}_{sol,w}(\mathbf{Q} = 0)$ is the difference in the solvation energy of the protein with and without the charged groups. All the terms in eq. (32) are obtained by the standard PDLD treatment except they are now scaled by a factor of the order of $1/\epsilon^{\alpha}_{in}$ (here, we use the fact that $1/\epsilon_w \simeq 0$). This scaling of the microscopic energy terms reduces the absolute value of the energy contribution and therefore increases the precision of the calculations.

At this point, one might wonder what value should be used for ϵ^{α}_{in}. In addressing this issue,[8] it is crucial to remember that ϵ^{α}_{in} is not the actual protein dielectric constant ϵ_{in} because we explicitly take into account the protein permanent dipoles in the $V_{Q\mu}$ term. In fact, one should realize that the dielectric constant represents all contributions that are *not* considered explicitly.[8] Thus, for example, if we were to consider explicitly all the polar contributions including the protein permanent and induced dipoles and their relaxations upon charge rearrangements (as done in the PDLD treatment), we should have used $\epsilon^{\alpha}_{in} = 1$. Here, we consider *implicitly* the protein-induced dipoles, some bound water molecules not included explicitly, and the reorganization of the protein permanent dipoles. Although it is possible to estimate the relevant ϵ^{α}_{in} for any system for microscopic simulation,[8] we will examine the results obtained for ϵ^{α}_{in} between 2 to 6 in this work.

Some Macroscopic Models

Although this article focuses on microscopic and semimicroscopic approaches, it is useful to consider here some macroscopic models. This is done to clarify and illustrate some of the problems associated with the macroscopic concepts. The first of these models involves the commonly used assumption that electrostatic energies in proteins can be assessed using $\epsilon = 2$ or, in our notation,

$$\Delta G = \frac{332}{2}\sum_{i<j}\frac{Q_iQ_j}{r_{ij}} = \frac{\Delta V_{QQ}}{2} \tag{33}$$

This treatment (referred to here as the "$\epsilon = 2$" model), which might seem quite reasonable, is equivalent to the implicit assumption that both the protein *and* the surrounding solvent can be considered as a nonpolar medium. The problems associated with this assumption will be illustrated in our comparative studies.

Another popular macroscopic model involves the use of a distance-dependent dielectric function $\epsilon(R) = R$ or, in our notation,

$$\Delta G = 332\sum_{i<j}\frac{Q_iQ_j}{r^2_{ij}} \tag{34}$$

The problems associated with these type of models (referred to here as the "$\epsilon = R$" models) were considered in ref. 3 and will also be illustrated in our comparative studies.

Finally, we will examine the performance of the Tanford–Kirkwood model.[41] This model considers the protein as a nonpolar medium surrounded by a high dielectric solvent. The problems associated with this and related models have been discussed before.[3,7] However, to clarify these problems it is useful to improve the model so that it correctly accounts for the shape of the protein (in this way, one can separate the problems associated with the assumption of a nonpolar protein from those associated with the shape of the protein). The improved model, referred to here as the Discretized Tanford–Kirkwood (DTK) model, considers the protein as a continuum with a dielectric of $\epsilon = 2$ while describing the solvent as a discretized continuum with a dielectric of $\epsilon = \epsilon_w$. The corresponding calculation can be done by using a finite element method, but it can also be done by using the PDLD/S approach,

omitting the $\Delta V_{Q\mu}$ term of eq. (32) while setting $\epsilon_{in}^{\alpha} = 2$ and $1/\epsilon_w \simeq 0$. This gives

$$\Delta G \simeq \frac{1}{2}\left[\left(\sum_i - \Delta G_{sol,\infty}^{i}\right) + \Delta G_{Qw} + V_{QQ}\right] \quad (35)$$

EXAMINING THE PERFORMANCE OF DIFFERENT MODELS

To assess the reliability of a given model for calculations of electrostatic energies in macromolecules, it is essential to use this model in systematic studies of different test cases. Such a validation study is in many respects more relevant than the formal justification of the different models. For example, a formally rigorous all-atom model can give entirely unreliable results when practical considerations dictate the use of a small cutoff radius.[31] In this section, we examine the various approaches of the method section by calculating the electrostatic free energies in different classes of problems, ranging from the simple cases of small molecules in water to much more challenging cases, such as ion pairs in protein interiors.

Solvation Energies of Small Molecules in Solution

Before examining the performance of different models for calculations of electrostatic energies in proteins, it is useful to examine such models by calculating the solvation energies of molecules and molecular ions in aqueous solutions. Such a test case is considered in this work, using both the PDLD and the FEP methods. The PDLD calculations involved the selection of 4 grid centers by randomly generating 40 centers, evaluating the corresponding energies by the fast noniterative Langevin model,[3] and keeping the 4 grids with the lowest energies. The final energy was then obtained by evaluating self-consistently the PDLD energies of the four grids [with 25 iterations of eq. (2)] and then averaging these energies. Note that a more consistent procedure would have required a Boltzmann average of the self-consistent energies of many randomly generated grids but the results of this expensive procedure are similar to those obtained using the current procedure. The calculations involved a grid size of 18 Å (i.e., $R_3 = 18$ Å).

The FEP calculations were performed using a radius of 14 Å for region 3 and 18 Å for region 4 and, of course, including the bulk energy of region 5. The temperature of the system was kept at 300 K and the FEP integration involved 11 mapping steps of 3 ps each with 2-fs time steps (see ref. 28 for a typical FEP study with the SCAAS model and for the corresponding error range, also see below the discussion of the convergence of our approach).

The results of our test cases are summarized in Table VI and depicted in Figure 6, which includes monoatomic ions of different sizes and charges, molecular ions, and polar molecules. As seen from the table, both the PDLD and FEP methods can give reliable results.

Table VI. Solvation energies of several ions and molecules in solution obtained by the PDLD and FEP methods.

System	PDLD	FEP	Obs.[a]	Ref.
Na^+	−100	−101	−100	64, 65
Ca^{2+}	−389	−377	−381	64
F^-	−108	−109	−106	52–55
Cl^-	−74	−76	−77	52–55
Br^-	−74	−72	−73	52–55
I^-	−60	−64	−63	64, 65
NH_4^+	−79	—	−78	68, 69
$N(CH_3)H_3^+$	−77	—	−74	48, 57–59
$N(C_2H_5)H_3^+$	−71	—	−66	69
$N(CH_3)_3H^+$	−62	—	−61	58–60
Asp^-	−83	−77	−80	3, 73
Ser^-	−85	−90	−90	60
His^+	−71	−64	−65	3
Ser	−9	−8	−7	74
His	−12	−10	−12	74
Asp	−10	−9	−9	74
Arg	−10	−14	−13	74
Asn	−14	−9	−12	74
Lys	−5	−5	−6	74
CH_4	2	—	2	68, 74
C_5H_{12}	1	—	2	68
C_6H_6	−1	—	−1	68
CH_3OH	−7	—	−5	72, 73
CH_3CH_2OH	−5	—	−5	73
C_6H_5OH	−8	—	−7	73
C_5H_5N	−4	—	−5	72
$C_5H_5NH^+$	−56	—	−54	69
$C_6H_5NH_2$	−7	—	−5	75

Calculated solvation energies (in kcal mol^{-1}) of several ions and molecules in solution. Some of the calculations were used to calibrate the PDLD and FEP parameters by adjusting the van der Waals parameters of both methods. The table emphasizes the PDLD results and only some FEP results are given. The solvation energies for the amino acids were obtained with the charges of the main chain atoms set to zeros and the corresponding observed values are taken relative to the solvation energy of glycine.

[a]The observed values listed here are average values of the experimental free energies from the sources indicated in the last column of the table. Many of the experimental results involved an error range of ±5 kcal mol^{-1}, in part due to the different possible interpretations and uncertainties associated with some parts of the thermodynamic cycles used for obtaining the experimental values.

Apparently, the reason for this success is that any model of polar solvents that *properly* treat the interplay between dipole–dipole and dipole–charge interactions can reproduce the main physics of charges in a uniform *homogeneous* polar medium. However, as demonstrated in different microscopic studies,[4,20,42] the essential step in building any reliable model is the *calibration* of the van der Waals radii of different atoms to reproduce observed solvation energies. The same is true for macroscopic approaches that give reasonable results with cali-

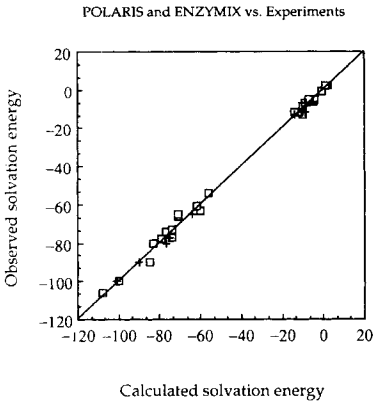


Figure 6. Plot of the observed and calculated solvation energies given in Table VI. The calculations are performed with the MOLARIS simulation package,[15] which includes the POLARIS PDLD calculations (□) and the ENZYMIX FEP calculations (+). A straight line corresponds to a perfect fit. This plot serves as a validation of the accuracy of our two microscopic models.

brated atomic cavities.[12] On the other hand, much more rigorous all-atom models can give unreliable results with *uncalibrated* van der Waals parameters, even when such parameters are deduced from quantum mechanical calculations.

Estimation of Transfer Free Energies in QSAR Studies

It is frequently assumed that the partition coefficients of molecules between nonpolar and aqueous media are correlated with their binding affinity to receptor sites (e.g., ref. 43). In fact, one of the primary guides in the design of biologically active molecules is the knowledge of their partition coefficients (P) and the corresponding transfer free energies. This subject is of major interest in studies of the so-called quantitative structure–activity relationships (QSARs). Many studies in this field focus on finding ways to estimate the relevant log P. The traditional approach is based upon trying to find an empirical correlation between the observed log P and different molecular properties.[43,44] However, one may try to calculate log P from a first principle by evaluating the solvation free energy of the given molecule in the different phases. Such calculations may be performed, for example, with the PDLD model and its field dependence hydrophobic correction term. A preliminary study of this type has been reported in ref. 16. Unfortunately, most of the transfer data in the literature are related to the partition between octanol and water rather than between nonpolar hydrocarbons and water. Considering the complexity of the solvation of molecules in octanol (which acts both as a polar and nonpolar solvent), we delayed to further works the attempt to create an LD model for octanol and introduce some empirical elements in our approach, that is, our current strategy of estimating log P for transfer of molecules from octanol to water is based upon taking the ΔG_{Qw} and $T\Delta S$ contributions of the solvation energies and scaling them by constants that give the best fit between the calculated and observed log P for a series of molecules. The results of our preliminary study are summarized in Figure 7 and the corresponding correlation coefficients are given in the figure legend.

As can be seen from Figure 7, our approach gives reasonable results considering the fact that we used only three adjustable parameters. The elucidation of the actual meaning of these correlation parameters would require further studies. It is important to note that the use of unreliable charges leads to poor results. Thus, we use a special approach in selecting the residual charges for the molecules included in our study, that is, the charges used were obtained by the AM1 method and then scaled by the ratio between the observed and calculated dipole moment ($Q_i = Q_i^{AM1}(\mu_{obs}/\mu_{AM1})$). This simple procedure resulted in a significant improvement of our model. Fianlly, it was found that changing the cutoff value ξ_{philic} of eq. (11) from 0.01 to 0.02 leads to better correlation, which allows us to use the same correlation coefficient for linear and cyclic molecules. However, a more systematic study of this issue is clearly needed.

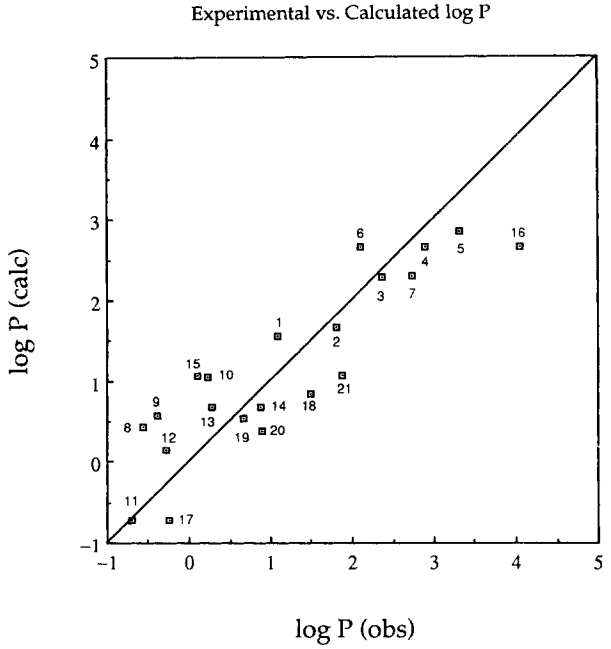


Figure 7. Correlation between the calculated and observed log P for transfer of molecules from octanol to water. The numbers in the figure correspond to the following molecules: methane (1), ethane (2), propane (3), butane (4), pentane (5), benzene (6), toluene (7), methylamine (8), dimethylamine (9), trimethylamine (10), methanol (11), ethanol (12), propanol (13) butanol (14), dimethylether (15), biphenyl (16), acetic acid (17), phenol (18), pyridine (19), aniline (20), and nitrobenzene (21). The experimental values (log P_{obs}) are taken from ref. 76, while the calculated values (log P_{calc}) are obtained by $2.3RT$ log $P = 0.34\Delta G_{QW} + T\Delta S - 3.65$ for $T = 300$ K.

Self-Energies of Charges in Solvated Proteins

The study of solvation energies of small molecules in solution is not the major challenge of this work, in particular because the van der Waals parameters or the atomic Born radii are adjustable parameters. A more serious challenge, which has been repeatedly addressed in our previous works,[2-4] is the evaluation of the self-energies of charges in protein interiors. A clear concept of such self-energies is crucial for understanding the energetics of many biologic processes, including redox problems, ion binding, and ion permeation in membrane channels.[4] In this section, we examine the ability of different models to reproduce the self-energies of single charged groups in several representative systems, which include the stringent test of the evaluation of the intrinsic pK_as of acidic groups in proteins, self-energy of the heme in cytochrome *c*, self-energy of calcium ions in a calcium binding protein, and barrier for the penetration of a sodium ion through the gramicidin A channel.

Although the self-energies of charges in proteins might seem abstract, they can be obtained uniquely from standard thermodynamic or kinetic properties (e.g., pK_as, redox potentials, conductance of ions, etc.). This point can be illustrated, for example, by considering the pK_as of ionizable groups in proteins. In this case, one uses a thermodynamic cycle of the type illustrated in Figure 8 (see ref. 45 for what is probably the earliest use of thermodynamic cycles in microscopic studies of proteins). This cycle and related considerations give[45,46]

$$pK_a^p = pK_a^w + \frac{1}{2.3RT}[\Delta\Delta G^p_{sol}(AH \longrightarrow A^-) - \Delta\Delta G^w_{sol}(AH \longrightarrow A^-)] \quad (36)$$

where the superscripts p and w denote that the corresponding quantities are evaluated in protein and water, respectively. $\Delta\Delta G_{sol}$ is the difference in solvation energy between the ionized (A^-) and un-ionized (AH) states of the acid. In many cases, the difference between the pK_as of an ionizable group in water and protein is mainly determined by the relative solvation energy (or self-energy) of the charged form of this group, $\Delta G_{sol}(A^-)$, in the given media. Thus, one may approximate eq. (36) by

$$\Delta pK_a^{w\to p} \simeq \frac{1}{2.3RT}\Delta\Delta G^{w\to p}_{sol}(A^-) \quad (37)$$

However, reliable studies should also consider the $\Delta\Delta G^{w\to p}_{sol}(AH)$ term. The same types of considerations are applied to redox problems,[18,47] ion channels,[48,49] and ion binding by proteins.[50] The main problem is to obtain reliable results for the *absolute* solvation energies of the charged groups in the complicated protein microenvironments. The performance of the three macroscopic models, as well as our three microscopic models, in these crucial tests is summarized in Table VII.

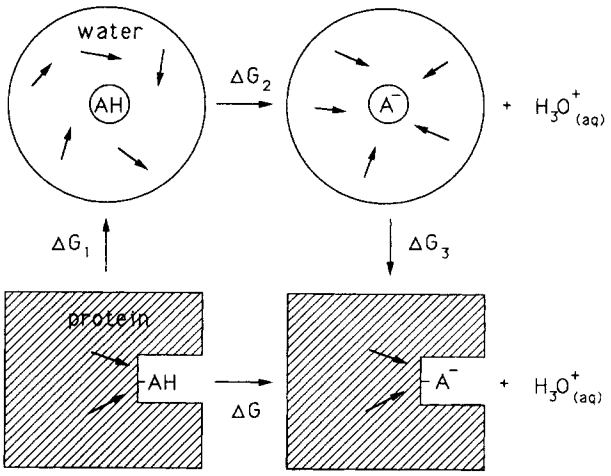


Figure 8. Thermodynamic cycle used to estimate the energetic of dissociation of an acidic group of a protein. The ΔG_i is given by $\Delta G_i = \Delta G^{w\to p}_{sol}(AH)$, $\Delta G_2 = 2.3RT(pK_a^w - pH)$, and $\Delta G_3 = \Delta G^{w\to p}_{sol}(A^-)$. The free energy of ionizing the group is given by $\Delta G = \Delta G_1 + \Delta G_2 + \Delta G_3$.

Table VII. Electrostatic energies of different ionized groups in proteins evaluated by different macroscopic and microscopic methods.

	Macroscopic			Semimicroscopic PDLD/S			Microscopic			
System model	$\epsilon = 2$	$\epsilon = R$	DTK	$\epsilon^a_{in} = 2$	$\epsilon^a_{in} = 4$	$\epsilon^a_{in} = 6$	PDLD	FEP	Obs.	Ref.
Bpti(Asp3)	—	—	6	−0.5	−0.2	−0.1	−3	−1	−1.3	51, 52
Bpti(Glu7)	—	—	15	3.5	1.7	0.8	1	2	−0.6	51, 52
Cytc(Heme)	—	—	13	12	6	4	10	6	7	47
Gramicidin A(Na^{+1})	—	—	25	5	2	1	2	6	5	48, 49
Calbindin(Ca^{+2})	−232	−175	−10	0.7	−2.5	−3.6	−6	—	−5	50

The electrostatic energies (in kcal mol^{-1}) for the DTK, PDLD/S, PDLD, and FEP methods are given relative to the corresponding solvation energies of the given group in water. The energies from the $\epsilon = 2$ and $\epsilon = R$ methods represent the interaction of the given group with other ionized groups of the protein. Note that for the $\epsilon = 2$ model the reference state (or the solvent around the protein) is a nonpolar continuum with $\epsilon = 2$. The observed values are taken from the indicated references. Note that the BPTI FEP results do not involve average over several initial configurations (as is done in Table VIII).

The bovine pancreatic trypsin inhibitor (BPTI) test case is indeed much harder than what may be deduced from the fact that observed pK_as of surface groups are usually similar to each other.[51,52] The apparently similar pK_as involve compensation of larger energy contributions of different nature (see ref. 10) and it is difficult to obtain such a compensation by microscopic approaches. In fact, the FEP model starts to give reasonable results only when large cut-off radii ($\geq$18 Å) are used. These calculations are expensive and the precision of the results are often poor. Fortunately, the recently introduced LRF method[17] increases substantially the speed of the all-atom FEP calculations while still giving the results obtained without any cutoff. The high performance of the LRF method allows us to use the computer power of the current workstations to average the results of FEP calculations over *different* sets of initial conditions. The results obtained by using this averaging procedure appear to be more accurate and precise than those obtained from a prolonged calculation started from a single set of initial conditions.[17] The self-energies of the four acidic groups of BPTI evaluated by the LRF all-atom model are given in Table VIII. These results are by far our best microscopic results for the BPTI test case or related pK_a calculations, that is, the present study appears to push the error range to the 1 kcal mol^{-1} range (although this small error range might still reflect an accidental success). On the other hand, our previous studies of the four-acids benchmark gave errors of the order of 5 kcal mol^{-1}.[4] Our early FEP study of the difference between Glu 7 and Asp 3 gave an error range of 1–2 kcal mol^{-1}, but this involved only two acids and might have also reflected an accidental success. Recent studies[53,54] that probably put less emphasis on the treatment of long-range interactions (in terms of the boundary conditions and cutoff treatments) obtained errors in the range of 5 kcal mol^{-1}.

It is important to comment here about the convergence of FEP calculations. Some recent studies[77,78] have given the impression that simulation lengths of several hundred picoseconds are required to obtain reasonable accuracy. Furthermore the long relaxation times of proteins and various polar solvents might indicate that rigorous convergence is beyond the range of any current computer. In fact, the early realization of this problem (e.g., ref. 3), led to the development of this simplified PDLD model. Fortunately, however, it was found in our studies that the convergence of electrostatic energies is much faster than that needed for evaluation of the energies associated with short-range forces or with hydrophobic interactions (which were studied in refs. 77, 78). The convergence of electrostatic energies is probably associated with the fact that different regions in configuration space give similar energies. Thus focusing on electrostatic energies is expected to give much faster convergence (e.g., see Fig. 11 of ref. 61). The problems associated with the slow convergence of steric energies can be partially overcome by estimating these contributions using the corresponding reorganization energy. It is also important to emphasize that the convergence of electrostatic energies depends drastically on the specific boundary conditions and on the proper treatment of long-range forces. The use of truncation methods with a cut-off radius and periodic boundary conditions do not provide the optimal strategy. In summary, we believe that the fact that the protein phase space is complicated and requires enormous time for proper samplings is the problem rather than the solution. Thus one should always look for the strat-

Table VIII. Self-energies of the acidic residues of BPTI evaluated with the FEP all-atom model augmented by the LRF method.

	Contribution	Asp 3	Glu 7	Glu 49	Asp 50
In water	ΔG^{w}_{sol}	−72.0 ± 1.1	−70.9 ± 2.0	−70.9 ± 2.0	−72.0 ± 1.1
In protein	$\Delta G^{p}_{Q_\mu}$	−14.2	−10.0	−33.3	−39.4
	$\Delta G^{w}_{Q_\mu}$	−48.2	−49.8	−27.8	−25.0
	$\Delta G^{p}_{Q_\alpha}$	−1.1	−2.3	−0.5	0.8
	ΔG_{bulk}	−9.4	−9.0	−9.4	−8.9
	ΔG^{p}_{sol}	−72.9 ± 1.6	−71.1 ± 3.0	−71.0 ± 1.6	−72.5 ± 2.0
	$\Delta\Delta G^{w\to p}_{sol}$	−0.9 (−1.3)	−0.2 (−0.6)	−0.1 (−0.3)	−0.5 (−0.3)

Average electrostatic energies (in kcal mol^{-1}) of charging the aspartic and glutamic acids in water (ΔG^{w}_{sol}) and the acidic residues of BPTI (ΔG^{p}_{sol}). The radii R_2, R_3, R_{4a}, and R_{4b} of the all-atom model are taken as 12, 12, 12, and 12 Å, respectively, for the water calculations, while 13, 18, 18, and 18 Å, respectively, for the protein calculations. Under these conditions, there are about 245 water molecules in the water cases and about 550–650 water molecules in the different protein cases. No Langevin dipole is used in all these cases. All reported quantities are obtained by averaging the results of eight independent 22 ps MD/FEP trajectories generated from different initial configurations of the systems. The table gives the different contributions to the total calculated free energies for the protein cases. $\Delta V^{p}_{Q_\mu}$ and $\Delta V^{w}_{Q_\mu}$ are the contributions from the interactions between the carboxy group of the relevant acidic residue with the protein and water atoms, respectively. ΔV_{Q_α} is the contribution from the interaction between the protein-induced dipoles with the permanent charges. ΔG_{bulk} is the contribution from the bulk. In parentheses are the observed values of $\Delta\Delta G^{w\to p}_{sol}$ derived from the experimental pK_as[51,52] using eq. (36).

Table IX. FEP/LRF self-energies of Asp 3 and Glu 7 of BPTI evaluated in the presence of another ionized acid.

Contribution	Asp 3	Asp 3 (Glu 7$^-$)	Glu 7	Glu 7 (Asp 3$^-$)
ΔG_{QQ}	—	40.0	—	32.7
$\Delta G^p_{Q\mu}$	−6.4	−10.1	−9.8	−13.4
$\Delta G^{qw}_{Q\mu}$	−50.8	−62.8	−47.9	−52.9
$\Delta G^p_{Q\alpha}$	−2.3	−3.4	−1.5	−4.9
ΔG_{lgvn}	−6.6	−15.9	−3.7	−10.4
ΔG_{bulk}	−6.8	−20.1	−6.7	−20.7
ΔG^p_{sol}	72.9 ± 3.2	72.3 ± 1.5	69.6 ± 1.2	69.6 ± 1.0

Average electrostatic energies (in kcal mol^{-1}) of charging the Asp 3, with and without Glu 7 ionized, and Glu 7, with and without Asp 3 ionized, in BPTI. The radii R_2, R_3, R_{4a}, and R_{4b} of the all-atom model are taken as 15, 15, 18, and 24 Å, respectively, for the protein calculations. Under these conditions, there are 311 water molecules and 629 Langevin dipoles in all the protein cases (because the designated center of each case is the same). Note that the calculations for Asp 3 and Glu 7 with the other group un-ionized are similar to those in Table VIII although the present calculations include a Langevin grid in region 4b. All the reported quantities are obtained by averaging the results of four independent 22-ps MD/FEP trajectories generated from different initial configurations of the systems. The table gives the different contributions to the total calculated free energies for the protein cases. ΔG_{QQ} and ΔG_{lgvn} are, respectively, the contributions from charge–charge interaction and the Langevin dipoles (see Table VIII for other notations). The observed shifts are not known exactly but an analysis of the behavior of the experimental titration curve indicates that the interaction between the ionized acids is less than 1 kcal mol^{-1}.

egies that give the optimal convergence for the property under consideration, rather than assume that longer simulation time will always give accurate results.

Although the evaluation of the self-energy of an individual acid by FEP methods is a major challenge, it appears that the evaluation of the self-energy of one acidic group in the presence of another ionized acid is an even greater challenge. Here, we meet this challenge by calculating the self-energy of Asp 3 of BPTI, with and without Glu 7 ionized, as well as that of Glu 7, with and without Asp 3 ionized. Our studies indicated that even larger cutoff radii (≥21 Å) are needed for stable results. These calculations would be expensive when the solvent effect within these cutoff radii is represented exclusively by explicit water molecules. A more affordable alternative is to use Langevin dipoles to replace the water molecules at the peripheral of the system, while still using the LRF method. The results of such calculations are given in Table IX, which show that the charge–charge interaction due to the presence of a charged group is compensated by other energy contributions to give self-energies similar to those without the charged group. This is encouraging because the analysis of the titration curve of BPTI indicates that the interaction between the acids is less than 1 kcal mol^{-1}, and reproducing such a small shift by microscopic simulations is far from simple.

The cytochrome test case is perhaps the least challenging. Here, the oxidized heme can be viewed as a charged sphere with a large radius. The "solvation" of such delocalized charges involves smaller forces and smaller energy contributions than those of ionized groups with more concentrate charge distribution and therefore involves smaller errors when evaluated by microscopic models.[47]

The self-energy of an ion at a site in a transmembrane channel reflects the activation barrier for ion permeation through that site. Thus, a correct evaluation of the self-energy of an ion in a channel is a prerequisite for obtaining the complete free energy profile of ion permeation. Here, we take the gramidicin A channel as a test case and calculate the self-energy of a sodium ion at the center of this channel. The result is indeed quite good considering the error range (20–30 kcal mol^{-1}) of other related studies (see ref. 48 for discussion), which appears to be due to the neglect of some key electrostatic contributions.[48]

The study of Ca^{2+} binding by calbindin is also challenging because the magnitudes of the different energy contributions to the self-energy are much larger than those considered in the previous cases, making it more difficult to obtain stable results. Here, we find that only approaches that involve reminimization of the apo-protein give reasonable results. Reminimizations of the protein structures allow for a correct evaluation of the protein dipolar reorganization energy [the λ of eq. (10)], which is a significant contribution to the electrostatic energy of multivalent ions (charging of monovalent ion often involves less protein dipolar reorganization) and must be included in the calculations. The results from the PDLD/S method implies that a larger value of ϵ^{α}_{in} ($\epsilon^{\alpha}_{in} = 8$) is needed to reproduce the observed result in this case as compared to other test cases (see other tables for comparisons).

With the exception of the calbindin case, the DTK model gives unreasonable results for single charges in proteins. Further, the $\epsilon = 2$ and $\epsilon = R$ models overestimate in a drastic way the self-energy of calcium ions in calbindin.

Ion Pairs in Solvated Proteins

Ion pairs play an important role in many biologic processes, including enzyme catalysis, subunit interactions, and photosynthesis. In fact, ion pair in-

teractions are the most commonly used example of electrostatic interaction in proteins. Thus, it is important to explore the abilities of different electrostatic models to reproduce the energetics of ion pairs. However, in doing so it is important to understand the fundamental difference between the use of the macroscopic Coulomb-type expression

$$\Delta G = 332 \frac{Q_i Q_j}{r_{ij}\epsilon_{eff}} = \frac{V_{QQ}(R)}{\epsilon_{eff}} \qquad (38)$$

and the corresponding microscopic expression[3]

$$\Delta G = 332 \frac{Q_i Q_j}{r_{ij}} + \Delta G_{sol}(r_{ij}) - \sum_i \Delta G^i_{sol}(\infty)$$
$$= V_{QQ}(R) + \Delta G_{sol}(R) - \sum_i \Delta G^i_{sol}(\infty) \qquad (39)$$

According to eq. (38), changing the interatomic distance of a pair of sodium chloride ions from r_1 to r_2 gives a free energy change $\Delta\Delta G = (332/\epsilon_{eff})(1/r_2 - 1/r_1)$. For large values of ϵ_{eff} (e.g., $\epsilon_{eff} = \epsilon_w$), $\Delta\Delta G \simeq 0$. However, to obtain the same result microscopically, using eq. (39), is much harder. In that case, $V_{QQ}(R)$ and $\Delta G_{sol}(R)$ must compensate each other and such a compensation involves detailed balance of two large quantities. Although the macroscopic ϵ_{eff} provides a powerful description of the compensation between V_{QQ} and ΔG_{sol} in homogeneous polar solvents, it *cannot* be deduced from macroscopic considerations. In fact, in cases of ion pairs in proteins one finds that ϵ_{eff} is drastically different from the small values frequently assigned to protein interiors (see the discussion in refs. 4 and 8). Thus, it is important to calculate ΔG via eq. (39) using microscopic or semimicroscopic approaches. Here, we examine such approaches by considering several test cases with increasing levels of complexity.

Ion Pairs in Water

Before considering ion pairs in proteins, it is instructive to consider charge–charge interactions in water. In doing so, it is important to realize that such interactions can be described accurately by using ϵ_{eff} between 40 to 80 in eq. (38),[3] which reflects an almost complete compensation between V_{QQ} and ΔG_{sol} upon changing the interatomic distance from infinity to the given r_{ij} (see Fig. 4). However, obtaining such a compensation from a microscopic model is a significant challenge even for a simple ion pair in water. Here, we consider such a test case by using the PDLD model to evaluate the energetics of Na^+ Cl^- in water. In this study, summarized in Table X, we examine the interaction of the ions, 12 Å apart, by evaluating the energy of charging the sodium with and without a Cl^- ion. The calculations with the standard PDLD model give a reasonable compensation so that despite a change of -27 kcal mol^{-1} in V_{QQ} the overall change $\Delta\Delta G$ is only 0.9 kcal mol^{-1}. This corresponds to an ϵ_{eff} of 31, while the actual interaction energy should correspond to an ϵ_{eff} of $\simeq$80 at such a large distance. Although the present result is satisfactory, we could obtain an even better result by a variant of our model that includes in μ_b of eq. (8) (which usually corresponds to the dipole of regions 1 and 2 of Fig. 2) the total dipole moment of the Langevin dipoles in region 3. The result of this improved model, summarized in the lower part of Table X, corresponds to an ϵ_{eff} of 46, quite impressive for a microscopic model.

Charge–Charge Interactions between Surface Groups in Proteins

Our second test case involves charge–charge interactions between surface groups in proteins. The strength of such interactions can be deduced uniquely by genetic engineering experiments and a good example is the effect of surface groups on the pK_a of His 64 of subtilisin.[55] Here, we consider the individual effect of two such surface groups, Asp 99 and Glu 156, which are over 10 Å away from His 64. This test case has been the subject of several recent macroscopic studies[56,57] that reproduced the observed trend with good precision. However, such precision is expected from macroscopic models as long as they deal with charge–charge interactions between surface groups (rather than self-energies) because such models simply scale the electrostatic interactions by a large dielectric constant. In fact,

Table X. Effect of a distant chloride ion on the self-energy of a sodium ion in water obtained by the PDLD method.

	PDLD						
System	ΔG_{QQ}	ΔG_{lgvn}	$\Delta G_{bulk'}$	ΔG_{bulk}	ΔG_{total}	$\Delta\Delta G$	ϵ_{eff}
Na^+	0.0	−94.2	−7.6	—	−101.8	—	—
Na^+ Cl^-	−27.7	−82.6	7.6	—	−102.7	−0.9	31
Na^+	0.0	−94.2	—	−7.7	−101.9	—	—
Na^+ Cl^-	−27.7	−82.6	—	7.8	−102.5	−0.6	46

The effect of a distant chloride ion (12 Å away) on the PDLD self-energy (in kcal mol^{-1}) of a sodium ion in water. The ΔG_{QQ}, ΔG_{lgvn}, $\Delta G_{bulk'}$, and ΔG_{bulk} are, respectively, the contributions from charge–charge interactions, the Langevin dipoles, the bulk energy while neglecting the contribution of the Langevin dipoles to the μ_b of eq. (8), and the bulk energy. The table demonstrates the large compensation of different energy contributions in microscopic approaches, which can lead to large relative errors. To gauge the success of our microscopic Langevin dipole model, the effective dielectric constant is estimated using $\epsilon_{eff} = -332/(R \cdot \Delta\Delta G)$, where R is the distance between the ion pair (see the text). The results show that our simplified solvent model can reproduce a large dielectric effect.

Table XI. PDLD calculations of the effect of removing surface-charged groups in subtilisin on the pK_a of His 64.

	PDLD							Obs.	
System	$\Delta G_{Q\mu}$	$\Delta G_{Q\alpha}$	ΔG_{lgvn}	ΔG_{bulk}	ΔG_{total}	$\Delta\Delta G_{N\to M}$	ΔpK_a	(ΔpK_a)	Ref.
Wild type	−74.6	−10.0	−8.4	35.0	−58.0	—	—	—	
D99S	−50.6	−14.2	−13.9	21.1	−57.7	0.3	−0.2	−0.4	55
E156S	−47.3	−16.3	−14.9	21.0	−57.5	0.5	−0.3	−0.3	55

The PDLD free energies (in kcal mol^{-1}) of charging (protonating) His 64 in the native subtilisin and two of its mutants, D99S and E156S. The $\Delta G_{Q\mu}$, ΔG_{ind}, ΔG_{lgvn}, and ΔG_{bulk} are, respectively, the contributions from charge–dipole interactions, induced dipole interactions, the Langevin dipoles, and the bulk. The pK_a shifts of His 64 due to either mutation are obtained using an expression similar to eq. (37), i.e., $\Delta pK_a = -\Delta\Delta G_{N\to M}/1.38$ at 300 K. The observed pK_a shifts correspond to an ionic strength of 0.005 M. The calculated results reproduce the observed weak interactions between His 64 and either surface-charged group.

the effective dielectric constant, ϵ_{eff}, for an ion pair on a planar surface with $\epsilon = \epsilon_{in}$ on one side and $\epsilon = \epsilon_w$ on the other is simply given by $(\epsilon_{in} + \epsilon_w)/2$, which gives an $\epsilon_{eff} \simeq 40$ (such a surface is a good approximation for the case when the protein radius is larger than the distance between the ions). Here, again it is much harder to reproduce the observed small interaction by microscopic than by macroscopic models. Our PDLD results for the interaction between His 64 and Asp 99 and Glu 156 are summarized in Table XI. The calculations are done by evaluating the self-energy of His 64 in the native protein and in the D99S and E156S mutants. For the case of the native protein, the unminimized X-ray structure is used for the calculations, while for the cases of the mutants the corresponding structures are obtained by a limited local relaxations of the native structure. In all three cases, the residue Asp 32, a crucial part of the catalytic triad, is also ionized. R_2 and R_3 (see Fig. 2) are, respectively, 18 and 22 Å. As seen from Table XI, we obtain encouraging results, demonstrating that the PDLD model is capable of *reproducing* large ϵ_{eff} as a result of a microscopic compensation between large energy contributions. It is important to mention that smaller R_2 or R_3 gives incorrect results (results not shown).

Charge–Charge Interactions in Protein Interiors

After considering two cases that involve small and rather trivial effective interactions, it is important to consider ion pairs in protein interiors that may involve much stronger interactions.[3,4] Here, we consider the test cases of the Ile 16–Asp 194 ion pair in trypsin[3,57,58] and the His 159–Cys 25 ion pair in papain.[3,59] The results of this study are summarized in Table XII. As seen from the table, we obtain from the DTK model unstable ion pairs because these ions would rather be in water than in the incorrectly assumed nonpolar protein sites (see the discussion in ref. 7). Or, in other words, the oppositely charged ions *repel* rather than attract each other. On the other hand, the $\epsilon = 2$ and $\epsilon = R$ models overestimate drastically the stabilization of these ion pairs. These types of problems do not appear in the PDLD, PDLD/S, and FEP models, which describe correctly the energies of ion pairs.

Ionic Strength Modulation of Charge–Charge Interactions in Proteins

Before leaving this section, it is instructive to examine the effect of ionic strength on charge–charge interactions in proteins. This effect has attracted major attention for many years, probably reflecting the fact that until the emergence of genetic engineering it was difficult to determine *experimentally* the magnitude of other electrostatic factors. Apparently, as pointed out repeatedly,[3,4] the effect of the ionic strength is much smaller than the electrostatic effect of the protein polarity. Nevertheless, it is clearly important to be able to evaluate the change in charge–charge interactions due to the surrounding ionic atmosphere. Such calculations can be performed by discretized continuum approaches (see ref. 6). Here, however, we use the approach described in the method section.

As a first step in the examination of our approach

Table XII. Energetics of ion pairs in proteins obtained by the macroscopic and microscopic methods.

				Semimicroscopic						
	Macroscopic			PDLD/S			Microscopic			
System model	$\epsilon = 2$	$\epsilon = R$	DTK	$\epsilon_{in}^{a} = 2$	$\epsilon_{in}^{a} = 4$	$\epsilon_{in}^{a} = 6$	PDLD	FEP	Obs.	Ref.
Trypsin(Ile16,Asp194)	−45	−25	30	2	1	0	−3	−4	−2.8	57, 58
Papain(His159,Cys25)	−40	−19	21	−6	−4	−3	−5	−5	−2.5	59

The free energies ion pairs in protein interiors are evaluated using the three macroscopic approaches, the PDLD/S method, the PDLD method, and the FEP method. The DTK, PDLD, PDLD/S, and FEP energies are given relative to the corresponding energies in water. The $\epsilon = 2$ and $\epsilon = R$ results are given relative to the corresponding energies at infinite separation (see also Table VII). The results demonstrate that ion pairs can be more stable in proteins than in water. Note that the macroscopic models give entirely incorrect results.

Table XIII. Effect of ionic strength on the energetics of an ion pair in water obtained by the PDLD method.

	Ionic strength		
Distance (Å)	$\Delta\Delta G$ (0.01 M)	$\Delta\Delta G$ (0.1 M)	$\Delta\Delta G$ (0.5 M)
3.0	0.1	0.5	1.2
5.0	0.1	0.4	0.9
10.0	0.0	0.2	0.4

The relative free energies (in kcal mol^{-1}) of a Na^+ Cl^- ion pair in water, with and without the ionic strength effect, calculated using the PDLD method. As seen from the table, the newly introduced term ΔG_{is} of the PDLD method is able to reproduce the observed trend, that is, the ionic strength modulation has a lesser effect on ion pair interactions with greater separation distance and the shielding effect is greater for higher ionic strength.

for calculating ionic strength effects, we evaluate the energetics of an ion pair in water at different ionic strength. The results of these calculations are summarized in Table XIII. As seen from the table, our simple treatment is able to reproduce the trend expected from the ionic environment, that is, the interaction between the ion pair decreases as the ionic strength increases and the same concentration of electrolytes has a lesser screening effect on the ion pair interaction at a greater interatomic distance.

Next, we consider the effect of the ionic strength on the ΔpK_a of His 64 of subtilisin upon the mutation of either Asp 99 or Glu 156 to a serine residue.[55] The results are given in Table XIV, which reproduces the observed effect reasonably well despite the simple nature of the present treatment. In fact, because ionic strength effects in proteins are small it is not clear if more sophisticated models are really justified (although microscopic studies of ionic strength effects present a computational challenge).

Electrostatic Free Energy in Enzyme Catalysis and Ligand Binding

Electrostatic energies play a major role in biologic specificity including catalysis and binding.[1-6,11] Thus, the ability to evaluate electrostatic energies in such processes is a prerequisite to any structure and function correlation. Here, we examine the performance of our models in several generic test cases of binding the catalysis.

Table XIV. Effect of ionic strength on the pK_a shifts of His 64 of subtilisin upon mutating Asp 99 or Glu 156 to a serine estimated by the PDLD method.

	ΔpK_a	
Ionic strength (M)	Asp99 → Ser	Glu156 → Ser
0.005	−0.2 (−0.4)	−0.3 (−0.4)
0.01	−0.2 (−0.4)	−0.3 (−0.4)
0.1	−0.2 (−0.3)	−0.3 (−0.3)
0.5	−0.1 (−0.1)	−0.2 —
1.0	−0.1 (−0.1)	−0.2 —

The effect of ionic strength on the pK_a shift of His 64 of subtilisin upon the removal of the distant surface-charged groups, Asp 99 or Glu 156, is evaluated using the PDLD method. The pK_a shifts are obtained using the expression given in Table XI. The results reproduce the observed small effect of ionic strength on electrostatic interactions in polar solvents. In parentheses are the observed pK_a shifts.[55]

Electrostatic Energies in Enzymatic Reactions

Many enzymatic reactions are characterized by a large *change* in the charges distribution of the reacting substrate between its ground state and transition state. The corresponding change in the electrostatic interaction between the enzyme and the substrate appears to account for a major part of the observed rate enhancement (relative to the relevant reference solution reaction). Here, we consider two such examples by evaluating the electrostatic contributions to the catalytic free energy of lysozyme[45] and trypsin.[60,61] We also consider the effect of the Asp 102–Ala mutation in subtilisin.[40,62] The performance of the different models in these test cases is summarized in Table XV. As seen from the table, all the macroscopic models give incorrect results for the trypsin and lysozyme cases, although these models seem to give reasonable results for the subtilisin case. The PDLD/S with ϵ_{in}^{α} between 4 and 6 accounts qualitatively for the observed effects in all cases, while the PDLD and FEP models appear to give more quantitative results in these specific test cases.

Calculations of Relative and Absolute Binding Energies

Electrostatic interactions play an important role in determining the strength and specificity of ligand binding processes. An excellent example is provided by the binding of phosphorylcholine (*P*-Col) analogs to the McPC603 antibody,[63] where the positive and negative ends of the ligands interact specifically with the corresponding negative and positive charges of the antibodies. Here, we take as a test case the absolute and relative binding energies of the analogs $(CH_3)_3N^+(CH_2)_nCOO^-$ with $n = 3$ and $n = 2$ (see ref. 31 for details). These analogs differ in length and are designated here by ♀ and ♂ for the $n = 3$ and $n = 2$ cases, respectively. In trying to reproduce the absolute binding energies one has to consider, in addition to the corresponding change in electrostatic energies, the contributions from hydrophobic energy $\Delta\Delta G_{hyd}$ of eq. (11) and reorganization energy λ of

Table XV. Reaction free energies of enzymatic reactions obtained by the macroscopic and microscopic methods.

	Macroscopic			Semimicroscopic PDLD/S			Microscopic			
System model	$\epsilon = 2$	$\epsilon = R$	DTK	$\epsilon_{in}^{a} = 2$	$\epsilon_{in}^{a} = 4$	$\epsilon_{in}^{a} = 6$	PDLD	FEP	Obs.	Ref.
Lysozyme(gs → ts)	−64	−43	−50	−27	−14	−9	−8	−5	−7	45
Trypsin(gs → ts)	−46	−22	20	−9.6	−5.7	−4.3	−12	−7	−7	60, 61
Subtilisin(D102A)	10	5	7	7	3	1	5	6	6	40, 62

The absolute reaction free energies (in kcal mol^{-1}) of the trypsin and lysozyme catalyzed reactions [going from the ground state (gs) to the transition state (ts)] are obtained with reference to the same reactions in water. In addition, the relative free energy of the reaction catalyzed by subtilisin and the D102A mutant is shown as an application of our methods to studies of the effect of mutations on the reaction energy profiles.

Eq. 12, as well as the entropic contribution associated with the configurational space available for the ligand in the binding site and in solution (this contribution is obtained by the rough estimate of ref. 31). The results of the calculations are summarized in Table XVI, and more details are given in ref. 31. As seen from the table, we obtain reasonable relative binding energies with the PDLD/S, PDLD, and the FEP methods, although the FEP method gives good results only when large cutoff radii are considered. The PDLD/S method is in particular encouraging as it gives good estimates of the absolute binding energies as well as the relative binding energies, which is a less stringent test. With the exception of the DTK model, the macroscopic models drastically overestimate the absolute binding energies, although the relative binding energies obtained with those models are in the right direction.

Considering the speed and stability of the PDLD/S method, we examine its performance on several more test cases that involve binding of different ligands to McPC603. The calculations involved in generation of 20 configurations during an MD simulation and the averaging over the corresponding PDLD/S results. The calculated binding energies and the corresponding IC_{50} values are compared to the relevant observed values in Table XVII. As seen from the table, we can reproduce qualitatively the observed trend. This indicates that the PDLD/S method should provide a powerful tool for estimating binding free energies and computer-aided drug design.

CONCLUDING REMARKS

This work presents a systematic study of the performance of several microscopic and semimicroscopic models for calculations of electrostatic energies in solutions and macromolecules. This includes the evaluation of the solvation energies of ions and molecules in water, self-energies of single charges in proteins, energetics of ion pairs in solutions and proteins, interaction between surface charges in proteins, and effect of ionic strength on such surface interactions, as well as electrostatic contributions to binding and catalysis in solvated proteins. Encouraging results are obtained by the microscopic and semimicroscopic models although a reliable treatment of the long-range effects and boundary conditions appears to be crucial for obtaining reliable results. The present study also illustrates the problems associated with some macroscopic models by using these models to evaluate the experimental observed energetics of charges in proteins. Approaches that seem to give reasonable result for charges on the surface of proteins give unreliable results for more stringent test cases.

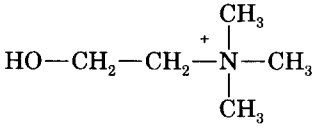
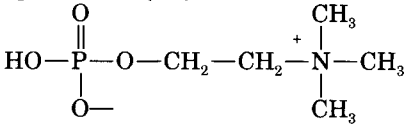
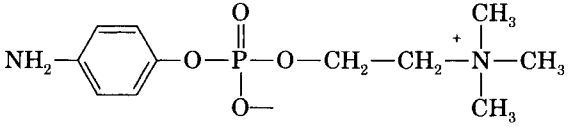
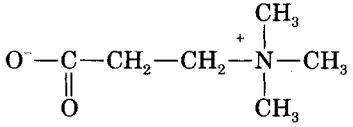
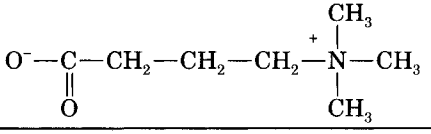
As argued repeatedly in our early studies (e.g., refs. 3 and 4), the effect of ionic strength is basically a second-order effect as much as the energetics of charges in protein interiors are concerned. Nevertheless, the present work evaluates the effect of the ionic strength on a macroscopic level using a grid of residual charges that satisfy the Boltzmann dis-

Table XVI. Absolute and relative binding free energies of phosphorylcholine analogs by McPC603 obtained by the macroscopic and microscopic methods.

	Macroscopic			Semimicroscopic PDLD/S			Microscopic			
System model	$\epsilon = 2$	$\epsilon = R$	DTK	$\epsilon_{in}^{a} = 2$	$\epsilon_{in}^{a} = 4$	$\epsilon_{in}^{a} = 6$	PDLD	FEP	Obs.	Ref.
McPC603(♀)	−64	−38	−5	−13.1	−7.3	−5.4	−8.6	−4.1	−7.1	[63]
McPC603(♂)	−57	−37	2	−8.1	−4.8	−3.7	−6.2	−2.7	−5.4	[63]
$\Delta\Delta G_{bind}^{\rightleftharpoons}$	7	1	7	5	2.5	1.7	2.4	1.4	1.7	[63]

Absolute and relative binding free energies (in kcal mol^{-1}) of $(CH_3)_3N^+(CH_2)_nCOO^-$ by murine myeloma protein McPC603. The analogs with $n = 3$ (♀) and $n = 2$ (♂) are examined. In this test case, the PDLD/S method gives particularly good results, demonstrating its potential use in computer-aid interactive docking procedures and binding studies.

Table XVII. PDLD/S estimates of the absolute and relative binding energies and the corresponding IC_{50}s for different ligands of McPC603.

Antibody	Ligand	ΔG^{calc}_{bind} ($\Delta\Delta G$)	ΔG^{obs}_{bind} ($\Delta\Delta G$)	$\log(IC_{50}/IC_{50ref})^{calc}$	$\log(IC_{50}/IC_{50ref})^{obs}$
McPC603	Choline (*C*) 	−3.6 (3.2)	−4.4 (3.5)	>5.3[a]	>5.8[a]
(PDLD/S estimate with $\epsilon = 4$)	Phosphocholine (*PC*) 	−6.7 (0.0)	−7.9 (0.0)	0.0[a]	0.0[a]
	p-Aminophenyl phosphocholine (*APPC*) 	−5.9 (0.9)	−6.5 (1.4)	1.4[a]	2.3[a]
	P-Col short analog (*S*) 	−7.5 (2.3)	−5.4 (1.7)	2.3[b]	2.8[b]
	P-Col long analog (*L*) 	−9.8 (0.0)	−7.1 (0.0)	0.0[b]	0.0[b]

Energies are in kcal mol^{-1}. IC_{50} represents the concentration of the given hapten required in inhibit McPC603 precipitation of the C polysaccharide of pneumococcus by 50%.

[a] $IC_{50ref} = IC_{50PC}$.

[b] $IC_{50ref} = IC_{50L}$.

tribution expected from the corresponding Coulomb's law with the dielectric constant of water. Basically, our philosophy is to first evaluate the relevant electrostatic interaction on a microscopic level for zero ionic strength and then estimate the ionic strength correction on a macroscopic level. This approach appears to be effective despite its simple nature.

Comparing the performance of our different microscopic and semimicroscopic approaches, we find it significant that the PDLD and PDLD/S methods are much faster than the FEP method while still giving reasonable results. In particular, the speed and simplicity of the PDLD/S method make it an effective strategy for estimating electrostatic energies in macromolecules. In fact, the much more rigorous FEP method can give unreliable results when the boundary conditions are treated incorrectly or when small cutoff radii are used.[31] However, the current boundary conditions of the ENZYMIX program and the treatment of the long-range effects by the LRF method[17] appear to give reliable results. Thus, considering the rapid increase in the available computer time we recommend the use of the entire PDLD, PDLD/S, and FEP methods in predicting electrostatic energies in macromolecules. When the three approaches give similar results, these results are likely to be reliable. However, when the results are different one has to examine the origin of the deviation and eventually get similar results.

Finally, it might be useful to comment on the conceptual advantage of using several strategies for studies of electrostatic energies in macromolecules. Basically, one deals with complicated systems where many traps may (and do) exist on the way to correct evaluation of the given energies. As depicted in Figure 9, one may consider the PDLD and PDLD/S methods a *ruler* that covers the range from the fully atomistic FEP method to the discretized continuum and other continuum approaches. Covering this range is

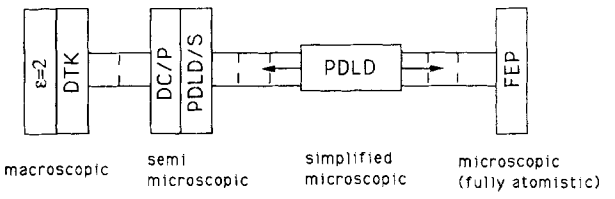


Figure 9. A simplified microscopic model such as the PDLD model provides a "ruler" that allows one to explore the range between the microscopic FEP models and the macroscopic models. At one end of our scale, we find the microscopic FEP method, based upon a fully atomistic representation. At the other end, we find the discretized continuum (DC) method, expected to give reasonable results only if the protein permanent dipoles are included explicitly [this model is referred to here as discretized continuum/polar (DC/P)]. The far end of the scale also includes the DTK model, which cannot give reliable results for charges in proteins. Between these macroscopic and microscopic extremes lie the PDLD and PDLD/S models.

useful not only in providing a fast alternative to the FEP method but also in protecting one from different pitfalls and providing error bounds for the given calculations, that is, we strongly recommend trying to estimate the electrostatic energy associated with a given problem by the three methods. When these methods give different results, it is likely that the problem is not treated properly by some, or all, the methods (e.g., incorrect boundary conditions or poor convergence). On the other hand, if the three methods give similar results we may trust these results and take the difference between them as a rough measure of the error range.

This work was supported by Grant GM24492 of the National Institutes of Health and Grant N00014-91-J-1318 of the Office of Naval Research.

References

1. M.F. Perutz, *Science*, **201,** 1187 (1978).
2. A. Warshel, *Proc. Natl. Acad. Sci. USA*, **75,** 5250 (1978).
3. A. Warshel and S.T. Russell, *Q. Rev. Biophys*, **17,** 283 (1984).
4. A. Warshel and J. Åqvist, *Annu. Rev. Biophys. Biophys. Chem.*, **20,** 267 (1991).
5. W.G.J. Hol, *Prog. Biophys. Mol. Biol.*, **45,** 149 (1985).
6. K.A. Sharp and B. Honig, *Annu. Rev. Biophys. Biophys. Chem.*, **19,** 301 (1990).
7. A. Warshel, S.T. Russell, A.K. Churg, *Proc. Natl. Acad. Sci. USA*, **81,** 4785 (1984).
8. G. King, F.S. Lee, and A. Warshel, *J. Chem. Phys.*, **95,** 4366 (1991).
9. A. Warshel and M. Levitt, *J. Mol. Biol.*, **103,** 227 (1976).
10. S.T. Russell and A. Warshel, *J. Mol. Biol.*, **185,** 389 (1985).
11. S.C. Harvey, *Proteins* **5,** 78 (1989).
12. A.A. Rashin, *J. Phys. Chem.* **94,** 1725 (1990).
13. M.E. Davis and A.J. McCammon, *Chem. Rev.*, **90,** 509 (1990).
14. A. Papazyan and M. Maroncelli, *J. Chem. Phys.*, **95,** 9219 (1991).
15. A. Warshel and S. Creighton, In *Computer Simulation of Biomolecular Systems*, W.F van Gunsteren and P.K. Weiner, Eds., ESCOM, Leiden, 1989, p. 120.
16. V. Luzhkov and A. Warshel, *J. Comp. Chem.*, **13,** 199 (1992).
17. F.S. Lee and A. Warshel, *J. Chem. Phys.* **97,** 3100 (1992).
18. R. Langen, G.D. Brayer, A.M. Berghuis, G. McLendon, F. Sherman, and A. Warshel, *J. Mol. Biol.* **224,** 589 (1992).
19. A. Warshel, *J. Phys. Chem.*, **83,** 1640 (1979).
20. A.K. Churg, R.M. Weiss, A. Warshel, and T. Takano, *J. Phys. Chem.*, **87,** 1683 (1983).
21. V. Frecer and S. Miertus, *Int. J. Quantum Chem.* **42,** 1449 (1992).
22. B.J. Klein and G.R. Pack, *Biopolymers*, **22,** 2331 (1983).
23. V. Daggett, P.A. Kollman, and I.D. Kuntz, *Biopolymers*, **31,** 285 (1991).
24. A. Warshel, *Photochem. Photobiol.*, **30,** 285 (1979).
25. A. Warshel, *Meth. Enzymol.*, **127,** 578 (1980).
26. P. Beroza, D.R. Fredkin, M.Y. Okamura, and G. Feher, *Proc. Natl. Acad. Sci. USA*, **88,** 5801 (1991).
27. A. Warshel, F. Sussman, and G. King, *Biochemistry*, **25,** 8368 (1986).
28. G. King and A. Warshel, *J. Chem. Phys.*, **91,** 3647 (1989).

29. A. Brunger, C.L. Brook, and M. Karplus, *Chem. Phys. Lett.*, **105,** 495 (1984).
30. C. Belch and M. Berkowitz, *Chem. Phys. Lett.*, **113,** 278 (1985).
31. F.S. Lee, Z.T. Chu, M. Bolger, and A. Warshel, *Prot. Eng.*, **5,** 215 (1992).
32. R.W. Zwanzig, *J. Chem. Phys.*, **22,** 1420 (1954).
33. J.P. Valleau and G.M. Torrie, *Modern Theoretical Chemistry,* Vol. 5, B.J. Berne, Ed., Plenum, New York, 1977, p. 169.
34. J.P. Postma, H.J.C. Berendsen, and J.R. Haak, *Faraday Symp. Chem. Soc.*, **17,** 55 (1982).
35. A. Warshel, *J. Phys. Chem.*, **86,** 2218 (1982).
36. A. Warshel, *Pontif. Acad. Sci. Scr. Var.*, **55,** 59 (1984).
37. C.F. Wong and J.A. McCammon, *J. Am. Chem. Soc.*, **108,** 3830 (1986).
38. S.N. Rao, U.C. Singh, P.A. Bash, and P.A. Kollman, *Nature,* **328,** 551 (1987).
39. D.L. Beveridge and F.M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.*, **18,** 431 (1989).
40. A. Warshel, G. Naray-Szabo, F. Sussman, and J.-K. Hwang, *Biochemistry,* **28,** 3629 (1989).
41. C. Tanford and J.G. Kirkwood, *J. Am. Chem. Soc.*, **79,** 5333 (1957).
42. J. Åqvist, *J. Phys. Chem.*, **94,** 8021 (1990).
43. A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71,** 555 (1971).
44. C. Hansch and A. Leo, *Substituent Constant for Correlation Analysis in Chemistry and Biology,* Wiley, New York, 1979.
45. A. Warshel, *Biochemistry,* **20,** 3167 (1981).
46. B.H. Honig, W.L. Hubbell, and R.F. Fleweling, *Annu. Rev. Biophys. Biophys. Chem.*, **15,** 163 (1986).
47. A.K. Churg and A. Warshel, *Biochemistry,* **25,** 1675 (1986).
48. J. Åqvist and A. Warshel, *Biophys. J.*, **56,** 171 (1989).
49. G. Eisenman and R. Horn, *J. Membr. Biol.*, **76,** 197 (1983).
50. S. Forsén, S. Linse, E. Thulin, B. Lindegard, S.R. Martin, P.M. Bayley, P. Brodin, and T. Grundström, *Eur. J. Biochem.*, **177,** 47 (1988).
51. R. Richarz and K. Wuthrich, *Biochemistry,* **17,** 2263 (1978).
52. K.L. March, D.G. Maskalick, R.D. England, S.H. Friend, and F.R.N. Gurd, *Biochemistry,* **21,** 5241 (1982).
53. J. Gao, K. Kuczera, B. Tidor, and M. Karplus, *Science,* **244,** 1069 (1989).
54. K.M. Merz, *J. Am. Chem. Soc.*, **113,** 3572 (1991).
55. A.J. Russell, P.G. Thomas, and A.R. Fersht, *J. Mol. Biol.*, **193,** 803 (1987).
56. M.K. Gilson and B.H. Honig, *Nature,* **330,** 84 (1987).
57. A.R. Fersht, *J. Mol. Biol.*, **64,** 497 (1972).
58. R. Huber and W. Bode, *Acc. Chem. Res.*, **11,** 114 (1978).
59. S.D. Lewis, F.A. Johnson, and J.A. Shafer, *Biochemistry,* **20,** 48 (1981).
60. A. Warshel and S.T. Russell, *J. Am. Chem. Soc.*, **108,** 6571 (1986).
61. A. Warshel, F. Sussman, and J.-K. Hwang, *J. Mol. Biol.*, **201,** 139 (1988).
62. P. Carter and J.A. Wells, *Nature,* **332,** 564 (1988).
63. N.M. Young, and M.A. Leon, *Immunochemistry,* **14,** 757 (1977).
64. H.L. Friedman and C.V. Krishann, In *Water—A Comprehensive Treatise,* Vol. 3, F. Franks, Ed., Plenum Press, New York, 1973, p. 54.
65. R.M. Noyes, *J. Am. Chem. Soc.*, **84,** 513 (1976).
66. D.R. Rossinsky, *Chem. Rev.*, **65,** 467 (1965).
67. W.A. Miller and D.W. Watts, *J. Am. Chem. Soc.*, **86,** 6051 (1967).
68. A. Ben-Naim and Y. Marcus, *J. Chem. Phys.*, **81,** 2016 (1984).
69. R.W. Taft, In *Progress in Physical Organic Chemistry,* Vol. 14, R.M. Taft, Ed., Wiley, New York, p. 247.
70. D.H. Aue, H.M. Webb, and M.T. Bowers, *J. Am. Chem. Soc.*, **98,** 318 (1984).
71. C.E. Klots, *J. Phys. Chem.*, **85,** 3585 (1984).
72. R.G. Pearson, *J. Am. Chem. Soc.*, **108,** 6109 (1986).
73. S. Cabani, P. Gianni, V. Mollica, and L. Lepori, *J. Solution Chem.*, **10,** 563 (1981).
74. P.A. Bash, V.C. Singh, R. Langridge, and P.A. Kollman, *Science,* **236,** 564 (1987).
75. C.J. Cramer and D.G. Truhlar, *J. Am. Chem. Soc.*, **113,** 8305 (1991).
76. N. Bodor, Z. Gabanyi, and C. Wong, *J. Am. Chem. Soc.*, **111,** 3783 (1989).
77. D.A. Pearlman and P.A. Kollman, *J. Chem. Phys.* **94,** 4532 (1991).
78. T.P. Straatsma and J.A. McCammon, *J. Chem. Phys.* **95,** 1175 (1991).