A General Treatment of Solvent Effects Based on Screened Coulomb Potentials

Sergio A. Hassan, Frank Guarnieri, and Ernest L. Mehler*

Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029

Received: November 3, 1999; In Final Form: March 7, 2000

A new implicit solvent model for Monte Carlo simulations of peptides and proteins is presented. The important features of the model are: electrostatics of polar solvation described by nonlinear, distance-dependent screened Coulomb potentials that are formally connected to classical dielectric theory of polar solvation; the self-energy is derived from the integral form of the Born equation; and the introduction of the Born radius of an atom embedded in a protein modeled as a dielectric continuum. The model was parametrized in the context of the CHARMM PAR22 force field by fitting to experimental solvation free energies of amino acid side chain analogues. Introduction of an intrinsic vacuum screening constant for each amino acid was required to obtain the correct screening behavior in solution. Solvation energies of several peptides in different conformations were calculated and compared with the corresponding quantities calculated with the Poisson−Boltzmann equation. The comparative results were highly correlated in all cases, and the correlation was independent of peptide length and charge. Moreover, the self-energy terms and the electrostatic interaction energies, separately calculated from the two methods, were also highly correlated. The computing time required for the model is about 5 times the vacuum calculation for small systems (∼100 atoms), but decreases to a factor of 3 as the system size increases. The development of a new hydrogen-bonding algorithm as well as initial results of using the new model for sequence-to-structure calculations on peptides is presented in the accompanying paper. The combination results in a complete continuum solvation model.

Introduction

The crucial role of water to regulate structure and function in proteins and other biologically active macromolecules is now well established.[1−4] Although the role of explicit water−protein interactions has been elucidated with the help of high-resolution X-ray crystallography,[5−7] the effect of bulk solvent on the electrostatic free energy of macromolecules and peptides has been much more difficult to describe. It is common, therefore, to perform theoretical studies using molecular dynamics (MD) calculations with an explicit representation of the solvent. Although these calculations have produced many interesting and important results,[8] the enormous computational requirements have precluded obtaining trajectories into relevant time regimes for realistic biological systems. Monte Carlo (MC) calculations with explicit solvent require small trial moves to obtain a reasonable acceptance rate, which makes a full exploration of configuration space computationally unfeasible.

Large reductions in computing time could be realized if the description of bulk solvent effects could be incorporated into the empirical potential function. Thus, efforts to develop "implicit solvent models" (ISM) have a history that parallels the development of MD and MC simulations (for a review see ref 9). An ISM must properly describe the two main effects of a polar solvent: the solvation of charges embedded in a polar medium, and the damping of the electrostatic interactions between charges of the solute due to polarization by the medium.[10,11] Many ISM are based on the solution of the Poisson−Boltzmann (PB) equation[12−18] or on the generalized Born (GB) approach.[19−26] Other models include integral equation or stochastic methods.[27,28] These methods assume that the protein is a uniform, low dielectric region, but some are still too slow for certain applications,[29,14] although improvements are being sought.[30] In contrast, the need for high dielectric screening to properly evaluate charge−charge interactions, even in the protein interior, and the regional dielectric variability that truly characterizes a particular protein has been continuously emphasized by Warshel and collaborators[10,11,31−33].

The GB approximation has been proposed as the basis to develop a computationally fast ISM. In its original form,[19] that is,

$$\Delta G_{\mathrm{solv}} = \frac{1}{2}(\epsilon^{-1} - 1) \sum_{\mathrm{i,j}} \frac{q_{\mathrm{i}} q_{\mathrm{j}}}{r_{\mathrm{ij}} + \delta_{\mathrm{ij}} r_{\mathrm{i}}}$$

where q_{i} is the net charge on particle i, r_{ij} is the distance between q_{i} and q_{j}, r_{i} is the Born radius for atom i, and δ_{ij} is the Kronecker delta. The GB was apparently a formal reexpression of a screened Coulomb potential (with screening ϵ in the dielectric and one in the vacuum) and the traditional Born term[34] for each atom of a molecule. Subsequent extensions in the functional form and parametrization of the GB were introduced to obtain better agreement with experimental data.[20] The GB approach has been applied with substantial success to solvation free energy calculations on small organic molecules, for which it was initially parametrized.[20] The approach has been developed for use with simulation techniques[22,20] and for modeling solvation in quantum chemical calculations.[21,35,36] The method is reasonably fast, and recent developments have provided more efficient formulas for calculating derivatives so that the method could be applied to MD simulations.[23,29,37] However, some recent studies suggest that when applied to proteins or small peptides the GB may not be as successful as it is for simple organic

* To whom correspondence should be addressed. Email: mehler@inka.mssm.edu.

10.1021/jp993895e CCC: $19.00
Published on Web 06/20/2000

molecules.[38−40] These results have suggested that optimization of the parameters based solely on solvation energies may not be sufficient for larger systems,[38] and that the GB may underscreen the interactions.[39] Very recently a new parametrization for proteins and nucleic acids has been reported,[41] but experience with this new parametrization is presently not extensive enough to determine if it resolves some of the apparent problems.

There is compelling experimental and theoretical evidence that dielectric screening is a nonlinear sigmoidal function of distance. Early studies of pK_a shifts in bifunctional organic acids and bases,[42−44] and dehydration studies on peptides and proteins[45−47] indicated that the observed screening of electrostatic interactions could be explained by assuming sigmoidal behavior.[48] The classic Lorentz−Debye−Sack (LDS) theory of polar liquids[49−52] leads to radially dependent permittivity profiles of sigmoidal form. When reaction field effects are incorporated into LDS theory, sigmoidal screening is enhanced.[53] LDS theory also could be used to calculate solvation energies of ions using the integral form of the Born equation.[34,53,54] Screened Coulomb potentials (SCP), which have been used for the calculation of electrostatic effects in proteins, can be formally derived from LDS theory thus demonstrating the fundamental nature of the SCP approach. The original form of the SCP[55] has recently been extended to enable calculation of the self-energy terms.[56,57]

Using the SCP, developed for the calculation of electrostatic properties in proteins, we now propose a complete ISM for use with MC and MD simulations. To test the validity of the ISM, the initial goal is the calculation of peptide structure from its sequence. In addition to a realistic representation of the solvent, attaining this goal requires an accurate description of the conformational potential energy landscape and a means of rapidly sampling the hyperdimensional torsion space to obtain a converged Boltzmann population of peptide structures. In the present work this is achieved by basing the ISM on the PAR22 force field[58] in CHARMM[59] that has been refined over many years, and the MC approach of conformational memories that was shown to completely sample conformation space in reasonable computing times.[60,61]

This paper is organized into four sections: In section I the formal connection between classic dielectric theory and screened Coulomb potentials is briefly reviewed.[9,53,54,62] In section II the formulation of the SCP-based ISM (SCP-ISM) is presented. General equations for the total electrostatic energy and total solvation free energy of macromolecules are derived and a new formulation of effective Born radii for atoms in a protein environment is introduced. The method is parametrized within the context of the PAR22 force field[58] of CHARMM.[59] In section III the method is used to calculate solvation free energies for different conformations of several peptides of different lengths with varying net charge, and the results are compared with calculations on the same systems using the finite difference approach for numerical integration of the PB (FDPB) equation. In addition, the results are analyzed in terms of the interaction and self-energy contributions to the total solvation energy. Finally, in section IV, CPU times required for the new solvation model are discussed and compared with explicit solvent calculations.

I. Formal Connection between Screened Coulomb Potentials and Classic Dielectric Theory of Lorentz, Debye, and Sachs

1. Radial Permittivity Profiles. According to LDS theory[49−53] the induced polarization in an element of a polar solvent at a point $\boldsymbol{r}$ from an electrostatic source (ion or dipole) is

$$P = \alpha E_l/\upsilon + \mu\langle f(\theta)\rangle/\upsilon \tag{1}$$

where α and υ are the polarizability and molecular volume, respectively, μ is the dipole moment, and $f(\theta)$ is the orienting factor, where the brackets indicate Boltzmann averaging over the orientations. The local (Lorentz) field, E_l, acting on a molecule in the solvent element is

$$E_l = E + (4\pi/3)P \tag{2}$$

where E is the external (macroscopic) field. For a point charge the orienting factor is $\cos\theta$, where θ is the angle between the solvent dipole and the vector connecting the point charge to the dipole center. Evaluating $\langle f(\theta)\rangle$[63] yields

$$P = \frac{\alpha}{\upsilon}E_l + \frac{\mu}{\upsilon}L(\mu E_l/kT) \tag{3}$$

where $L(x) = \coth(x) - 1/x$ is the Langevin function, which can be approximated by $x/3$ for small x (low field). From the low-field approximation and the relation

$$E_l = (\epsilon + 2)E/3 \tag{4}$$

between the local and macroscopic fields, Debye derived his expression relating the dipole moment to the molecular volume, the high-frequency permittivity, ϵ_∞, and the static permittivity, ϵ_s.[49] For strong fields eq 3 must be used, but it is assumed that only the orientation of the dipole moment is affected; its magnitude is the same as in the low-field limit. For a particle with charge Q in a polar or polarizable medium the macroscopic field is $E(r) = Q/\epsilon(r)r^2$, so that

$$E_l(r) = \frac{\epsilon(r) + 2}{3}\frac{Q}{\epsilon(r)r^2} \tag{5}$$

From this expression and eq 3 the radial dielectric permittivity is expressed as

$$\epsilon(r) = 1 + \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2}[\epsilon(r) + 2] + \frac{4\pi\epsilon(r)\mu r^2}{Q\upsilon}L\left\{\frac{Q\mu[\epsilon(r) + 2]}{3r^2kT\epsilon(r)}\right\} \tag{6}$$

where the Lorenz−Lorentz relation has been used in the second term on the right hand side of eq 6. Equation 6 was first derived by Bucher and Porter[54] and subsequently by Ehrenson,[53] who also gave an expression for $\epsilon(r)$ in the case of dipole field sources.

In deriving eq 6 it was assumed that the directing field is the full Lorentz field acting at the position of the molecule, which is correct for spherically symmetric cases. For dipole fields (and for asymmetric or fluctuating ion fields) Onsager showed that part of the local field acting on the dipole is due to its own field acting on the surrounding solvent (the reaction field, R).[64] Because the instantaneous direction of the reaction field is the same as μ, it does not contribute to the orienting torque, and for these cases the directing field is given by

$$E_d = E_l - R\langle\cos\theta\rangle \tag{7}$$

where $R\langle\cos\theta\rangle$ is the average contribution of the reaction field to the local field. Equation 7 was derived by Böttcher, and allowed him to present a simple derivation of Onsager's result.[65] From eq 7 and assuming the low field approximation, Böttcher obtained the expression

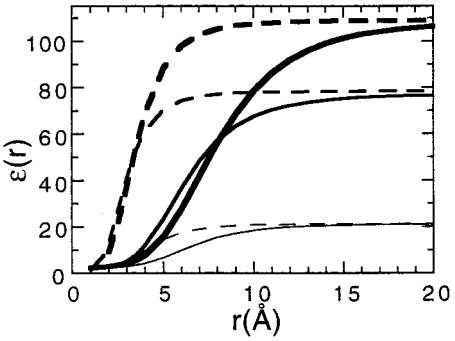

Figure 1. Permittivity profiles for formamide (thick lines), water (medium lines), and acetone (thin lines). Solid lines: $\epsilon(r)$ calculated from eq 6; dashed lines: $\epsilon(r)$ calculated from eq 6′ including the reaction field corrections. The reaction field was calculated using eq 22 in ref 53.

$$\langle \cos\theta \rangle = \frac{\mu}{3kT + \mu R} E_I \tag{8}$$

Using this result Ehrenson obtained a reaction-field-corrected form of the radially dependent permittivity profile $\epsilon(r)$, which for ion field sources has the form[53]

$$\epsilon(r) = 1 + \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2}[\epsilon(r) + 2] + \frac{4\pi\epsilon(r)\mu r^2}{Q\upsilon} L\left\{\frac{Q\mu[\epsilon(r) + 2]}{r^2\epsilon(r)(3kT + \mu R)}\right\} \tag{6′}$$

It is important to realize that eqs 6 and 6′ are valid for point charge field sources only. For dipole field sources Ehrenson[53] obtained analogous, but different, expressions for $\epsilon(r)$. Differences in the form of the screening from ionic and dipolar field sources was discussed earlier by Pollock et al.,[66] and in the context of protein electrostatics by Warshel and collaborators.[67] Finally it is noted that for asymmetric, fluctuating point charge sources (e.g., as encountered in proteins), the reaction field may not be of the same magnitude as for dipolar solutes.

The radial permittivities expressed by eqs 6 and 6′ depend only on physically measurable quantities. Both equations have been solved iteratively to obtain the permittivity profiles for acetone, water, and formamide, with the results presented in Figure 1. The most apparent effect of the reaction field correction is the compression of the profile to a much smaller range of radial values. Thus, for all three substances ϵ_{cdb} (Ehrenson form of the reaction-field-corrected permittivity) has nearly reached its asymptotic value at $r = 5$ Å, whereas for the uncorrected profiles this value is not reached until about 15 Å. For small values of r (<3 Å) the behavior of $\epsilon_{cdb}(r)$ is also very different from that of $\epsilon_{db}(r)$. These differences arise from the reduction in the directing field due to the reaction field contribution to the local field, which results in smaller solvent saturation effects.[53] In contrast, both profiles go to the same limiting value (see Table 1), provided that the correct value for the magnitude of the dipole moment is used, that is, μ_{db} for $\epsilon_{db}(r)$ and μ_{cdb} for $\epsilon_{cdb}(r)$, obtained from the low-field-limiting expressions of eqs 6 and 6′.[53] Table 1 gives the calculated limiting values of the static permittivity. The agreement with experimental values (ϵ_s in Table 1) is obvious. The dipole moments calculated in the Debye formalism for these three solvents are always smaller than the gas-phase-measured moments. Note also that the corrected moments calculated with Onsager's assumption ($\epsilon_i = 1$)[64] are a little larger than the values calculated with Ehrenson's assumption ($\epsilon_i = \epsilon_\infty$).[53]

2. Screened Coulomb Potentials. Effective electrostatic screening was first introduced to rationalize the observed ratios of the second to first dissociation constants of bifunctional organic acids and bases.[68] The interaction between the charged functional groups could be calculated from the relation $\Delta w = q_1q_2/\epsilon_s R = 2.303kT\Delta p\mathrm{K}$.[69] Here R is the separation of the charges q_1 and q_2, located at the titration sites, ϵ_s is the solvent's dielectric constant, and k is the Boltzmann constant. This formula was used to estimate R, and gave reasonable results for long, straight-chain acids, but for short chains the value of R was much too small. Defining an effective screening,[68] that is, $D_e = q_1q_2/\Delta wR$, resulted in a reasonably smooth variation of D_e,[42−44] producing a sigmoidal form reaching its asymptotic limit around 20 Å.[70] To represent these results in an analytically convenient way that is in agreement with LDS theory, the effective screening was defined by a differential equation that describes sigmoidal forms,[55] that is,

$$\mathrm{d}D(r)/\mathrm{d}r = \lambda\,(1 + D)\,(\epsilon_s - D) \tag{9}$$

with the solution

$$D(r) = (\epsilon_s + 1)/\{1 + k\exp[-\lambda(\epsilon_s + 1)r]\} - 1 \tag{10}$$

where k is a constant of integration, ϵ_s is the dielectric constant of the solvent, and λ is the parameter that controls the rate of change of $D(r)$. This form of Coulombic screening will be used for constructing the implicit solvent model. For this purpose it is reasonable and convenient to set $D(0) = 1$ by taking $k = (\epsilon_s - 1)/2$. Thus $D(r)$ is defined by the single parameter, λ (see below).

For a particle with charge Q in a dielectric medium, the potential consists of charge and polarization contributions[71] and can be expressed as $\phi = Q/D(r)r$, where $D(r)$ incorporates all the effects from the dielectric media around the charge.[33] Although this expression neglects asymmetry effects, it will be demonstrated by application that, at least to a first approximation, these effects are small. From the relation $\boldsymbol{E} = -\nabla\phi$ and spherical symmetry it is easily shown that[9,62]

$$\epsilon(r) = D(r)\left[1 + \frac{r}{D}\frac{\mathrm{d}D}{\mathrm{d}r}\right]^{-1} \tag{11}$$

Equation 11 is the formal connection between the screened Coulomb potential of eq 10 and the radially dependent dielectric permittivity profiles of eqs 6 and 6′ derived from classic LDS

TABLE 1: Input Parameters to Calculate Radial Permittivity Profiles and the Limiting Values of Dipole Moments and Dielectric Permittivities[a]

solvent (temp)	vol (Å^3)	ϵ_∞[b]	μ_g	μ_{db}	μ_{cdb}	μ_{ons}	ϵ_{db}	ϵ_{cdb}	ϵ_s[c]
water (25 °C)	30.00	1.78	1.87	0.82	2.77	3.10	77.19	78.33	78.39
acetone (20 °C)	122.00	1.85	2.88	1.52	2.79	3.11	21.16	21.19	21.20
formamide (20 °C)	65.99	2.09	3.73	1.16	4.27	4.69	108.5	109.0	109.0

[a] Subscripts: g, gas phase; db, Debye theory; cdb, reaction-field-corrected Debye theory with $\epsilon_i = \epsilon_\infty$; ons, reaction field corrected with $\epsilon_i = 1$, where ϵ_i is the permittivity of the (virtual) cavity surrounding the charge. [b] High-frequency permittivity. [c] Experimental static permittivities (taken from the Handbook of Physics and Chemistry, 97th ed.).

theory. From eq 11 it is apparent that if $D(r)$ has sigmoidal form, so does $\epsilon(r)$, and the converse is also true. Thus, if $\epsilon(r)$ is known, $D(r)$ can be obtained by integrating eq 11, but a simpler procedure was developed earlier.[72]

3. Electrostatic Energy. From the above discussion of LDS theory and screened Coulomb potentials, it is now possible to obtain an expression for the electrostatic energy of a distribution of point charges, q_i, in a dielectric medium, expressed in terms of the screening function $D(r)$. The electric field for the collection of charges is given by

$$E(\boldsymbol{r}) = \sum_i \boldsymbol{r}_i' \frac{q_i}{\epsilon(r_i) r_i^2} = \sum_i E_i(\boldsymbol{r}) \tag{12}$$

where $r_i = |\boldsymbol{r} - \boldsymbol{r}_i|$ and $\boldsymbol{r}_i' = (\boldsymbol{r} - \boldsymbol{r}_i)/r_i$. From eqs 11 and 12 the electrostatic energy can be expressed as

$$\frac{1}{8\pi}\int \boldsymbol{E}\cdot\boldsymbol{D}\, d\upsilon = \frac{1}{8\pi}\sum_{i\neq j}\int \boldsymbol{E}_i\cdot\boldsymbol{D}_j d\upsilon + \frac{1}{8\pi}\sum_i \int \boldsymbol{E}_i\cdot\boldsymbol{D}_i d\upsilon$$

$$= \frac{1}{2}\sum_{i\neq j}\frac{q_i q_j}{D(r_{ij}) r_{ij}} + \frac{1}{2}\sum_i \frac{q_i^2}{D(R_i) R_i} \tag{13}$$

where $\boldsymbol{D}$ is the displacement. The first sum on the right hand side of eq 13 is obtained using $\boldsymbol{E} = -\nabla\phi$, eq 11, and the divergence theorem.[63] The second sum is obtained by inserting eq 11 into the integral of the diagonal term and direct integration.[72] In eq 13 $D(r_{ij})$ and $D(R_i)$ are both defined by eqs 9−11, where r_{ij} is the distance between the (partial) point charges q_i and q_j, whereas R_i represents a Born-like radius to be defined below. In this formulation two points are of importance: first, the sums in eq 13 run over *all* the partial charges defining the system, regardless of the net charge on the group to which a given atom belongs; and second, the function D and the radii R_i completely describe the character of the system. From eq 13 it is clear that the formulation of the electrostatic problem in terms of radially dependent dielectric permittivities permits a considerable simplification because it is no longer necessary to define the solute interior with permittivity ϵ_i, the continuum region with permittivity ϵ_s, and a discontinuous boundary between them. Thus the arguments concerning the correct value of ϵ_i and where to place the boundary between the solute and solvent disappear. This is an important simplification because of the well-known sensitivity of electrostatic calculations to both of these quantities.[15,29,66] In contrast, $\epsilon(r)$ is well-defined from LDS theory. With eq 13 we are now in a position to construct the SCP-ISM.

II. Formulation of the SCP-ISM

1. Total Electrostatic Energy and Solvation Free Energy. To obtain an expression for the total electrostatic energy of a macromolecule immersed in a dielectric medium, it is most convenient to construct a thermodynamic cycle that allows the system to reach its final configuration by following two alternative paths, paths I and II defined in Figure 2. This approach is quite similar to that proposed by other authors,[10,23,73,74] but to illustrate the new features of the ISM it is convenient to provide a brief recapitulation here. It is assumed that in its final configuration (in which the energy will be evaluated) the molecule is formed by a set of N partial charges $\{q_i\}$. At the initial stage (1) the particles are placed at infinite separation, in a vacuum, and each brought to its charge, q_i.

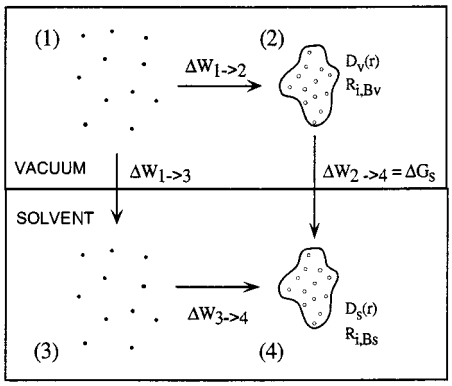

Figure 2. Thermodynamic cycle used to calculate the different contributions to the electrostatic energy of a solvated macromolecule. (1) Particles comprising the macromolecule at infinite separation; (2) macromolecule in a vacuum in its final configuration characterized by screening function $D_v(r)$ and Born radii $R_{i,Bv}$; (3) particles in solvent at infinite separation characterized by ϵ_s and $R_{i,w}$; (4) macromolecule in solvent in its final configuration characterized by screening function $D_s(r)$ and Born radii $R_{i,Bs}$. Path I is the process (1) → (2) → (4) and path II is (1) → (3) → (4).

Following path I the particles are then brought from infinite separation to their final relative positions in the molecule (2) and subsequently the molecule is transferred to the solvent (4). The total electrostatic energy, E_T, is

$$E_T = \Delta W_{1\to 2} + \Delta W_{2\to 4} \tag{14}$$

where $\Delta W_{2\to 4}$ is the solvation free energy, $\Delta G_s = \Delta G_s^{pol} + \Delta G_s^{nonpol}$. Here the first term is the electrostatic contribution, whereas the second term comprises all the remaining effects (see below). The energy required to construct the molecule in a vacuum is given by

$$\Delta W_{1\to 2} = W_{12,elec} + W_{2,self} - W_{1,self} \tag{15}$$

$W_{12,elec}$ is the Coulomb electrostatic energy of the molecule in state (2), given by

$$W_{12,elec} = \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j}{D_v(r_{ij}) r_{ij}} \tag{16}$$

The screening function, $D_v(r)$, accounts for the average effect of all the possible screening mechanisms within the macromolecule *in a vacuum*.

The process defined by the change in the self-energy, $W_{2,self} - W_{1,self}$, is a typical solvation problem where the molecule is constructed by adding atoms until the final structure is attained. Assuming the adequacy of a macroscopic treatment the change in self-energy can be treated in the framework of the Born theory of ion solvation. Therefore,

$$W_{2,self} - W_{1,self} = \frac{1}{2}\sum_i^N \frac{q_i^2}{R_{i,Bv}}\left[\frac{1}{D_v(R_{i,Bv})} - 1\right] \tag{17}$$

where $R_{i,Bv}$ is the Born radius of atom i to be defined in the process $1 \to 2$, that is, when the particle is brought from vacuum into the "solvent" composed of the protein *in a vacuum*.

Next the molecule is constructed by following path II in Figure 2. Starting from (1) each particle is solvated independently at infinite separation, state (3), and then the molecule is constructed by bringing the particles to their final configuration (4). The total electrostatic energy of the system is now given

by $E_T = \Delta W_{1\rightarrow 3} + \Delta W_{3\rightarrow 4}$. The term $\Delta W_{3\rightarrow 4}$ is analogous to $\Delta W_{1\rightarrow 2}$ defined in eq 15 except that the process is now embedded in the solvent. The presence of the surrounding solvent can be expressed by (a) replacing $D_v(r)$ by $D_s(r)$ in eqs 16 and 17, where the function $D_s(r)$ accounts for all the screening mechanisms of the system when the macromolecule is immersed in the given environment, and (b) replacing $R_{i,Bv}$ by $R_{i,Bs}$ in eq 17 where $R_{i,Bs}$ is the Born radius of atom i to be defined in the process $1 \rightarrow 4$, that is, when the particle is brought from vacuum into the protein *embedded in surrounding solvent*. It is clear, for example, that $D_s(r)$ in a protein is different from its analogue in pure water.

From path II and the Born model, an explicit expression of the total electrostatic energy can be obtained, that is,

$$E_T = \frac{1}{2}\sum_{i\neq j}^{N} \frac{q_i q_j}{D_s(r_{ij}) r_{ij}} + \frac{1}{2}\sum_{i}^{N} \frac{q_i^2}{R_{i,Bs}}\left[\frac{1}{D_s(R_{i,Bs})} - 1\right] \quad (18)$$

From eq 14 and eqs 15−18 the polar part of the solvation free energy is given by

$$\Delta G_s^{pol} = \frac{1}{2}\sum_{i\neq j}^{N} \frac{q_i q_j}{r_{ij}}\left[\frac{1}{D_s(r_{ij})} - \frac{1}{D_v(r_{ij})}\right] + \frac{1}{2}\sum_{i}^{N} q_i^2\left\{\frac{1}{R_{i,Bs}}\left[\frac{1}{D_s(R_{i,Bs})} - 1\right] - \frac{1}{R_{i,Bv}}\left[\frac{1}{D_v(R_{i,Bv})} - 1\right]\right\} \quad (19)$$

For a single ion, taking $D_s = \epsilon =$ constant and $D_v = 1$, eq 19 reduces to the standard form of the polarization energy in the Born approach.[34] In any case D_v is expected to be much smaller than D_s.[67]

The nonpolar contribution of $\Delta W_{2\rightarrow 4}$ in eq 14 is composed of the cavity term and the van der Waals term and is usually written in terms of the total solvent-accessible surface area (SASA) in the form[75]

$$\Delta G_s^{nonpol} = a + b\sum_{i=1}^{N} SASA_i \quad (20)$$

where $SASA_i$ is the solvent-accessible surface area of the individual particles, and a and b are parameters usually evaluated by fitting solvation free energies of hydrocarbons, for which $\Delta G_s^{pol} \approx 0$. To calculate the parameters a and b in eq 20 the SASA of each of the four analogues of Ala, Ile, Leu, and Val was calculated using CHARMM.[59] The least-squares fitting of the solvation energy of the four nonpolar analogues yielded $a = 1.0725$ kcal mol^{-1} and $b = 0.0042$ kcal Å^{-2} mol^{-1}, in close agreement with previous evaluations of these two parameters.[21,76,77] The total solvation free energy is evaluated from eqs 19 and 20, and the total electrostatic energy is evaluated with eq 18. Thus the SCP-ISM is completely defined.

2. Born Radius in a Protein Environment. The process of transferring a particle from the vacuum to the macromolecule can be considered as a typical solvation problem, provided that the system can be characterized by an effective dielectric screening function, $D(r)$, and appropriate Born radii can be defined for each atom in the molecule. Thus, in state (2) the system was characterized by a screening function, $D_v(r)$, and the Born radii of the atoms are defined by $R_{i,Bv}$, whereas in state (4) the system is characterized by a screening function $D_s(r)$ and Born radii $R_{i,Bs}$. A typical situation for an atom i partially exposed to solvent is schematically represented in Figure 3. The standard Born radius of a charged particle depends on the characteristics of the solvent, as discussed by other authors,[3,10,78,79] whereas the effective Born radius is also expected to depend on the degree of exposure to the solvent. Therefore we propose for the Born radii $R_{i,Bs}$ the following linear relation

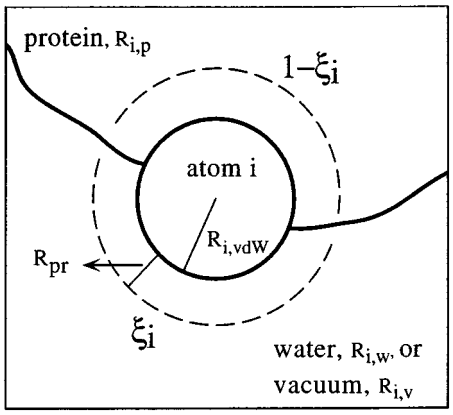

Figure 3. Schematic diagram showing the degree of exposure of an atom i to the solvent (ξ_i) and to the protein interior ($1 - \xi_i$), used to define the effective Born radius of an atom in the protein environment (cf. eqs 21 and 22). The Born radius of the atom in the protein is $R_{i,p}$; in the aqueous (vacuum) phase is $R_{i,w}$ ($R_{i,v}$).

$$R_{i,Bs} = R_{i,w}\xi_i + R_{i,p}(1 - \xi_i) \quad (21)$$

where $R_{i,w}$ is the Born radius of atom i in bulk solvent (the same radius that would be defined in the solvation process $1 \rightarrow 3$ of individual atoms) and $R_{i,p}$ is the Born radius of an atom immersed in an ideal bulk solvent composed of the protein interior considered as a dielectric continuum. The weighting factor ξ_i is the degree of exposure of the atom to the (real) solvent, and is defined by $\xi_i = SASA_i/4\pi(R_{vdW} + R_{pr})^2$. Here R_{vdW} is the van der Waals radius of the atom, R_{pr} is the probe radius, and $(1 - \xi_i)$ is the fraction of the atom's surface buried in the protein. Similarly we define the Born radius $R_{i,Bv}$ by

$$R_{i,Bv} = R_{i,v}\xi_i + R_{i,p}(1 - \xi_i) \quad (22)$$

where $R_{i,v}$ is the characteristic radius of the atom. Next we need to evaluate $R_{i,v}$, $R_{i,w}$, and $R_{i,p}$.

The Born radius of a solvated ion can be thought of as the radius of the cavity formed by the ion in the solvent; it can be interpreted as the distance from the center of the ion to the point where the solvent medium around the ion actually "starts".[78] The original ideas in this sense go back to the work of Latimer and Rodebush[80] in their studies of solvation of alkali metal and halide ions. These authors proposed that the Born radius of an ion is given by its ionic radius plus an extension that represents the enlargement of the cavity with respect to the "surface" of the atom as defined by the ionic radius. Because of the asymmetric location of the center of the permanent dipole moment in a water molecule, the extension of the ionic radius will be different for hydrated anions and hydrated cations.[54,80] Here, we will use this same idea of extending a characteristic radius of the atom to define the Born radii $R_{i,v}$, $R_{i,w}$, and $R_{i,p}$ in eqs 21 and 22.

Because the atoms in a macromolecule are not isolated ions but atoms with partial charges covalently bound, we propose to use the covalent radius, $R_{i,COV}$, instead of the ionic radius, as the fundamental quantity to define the characteristic size of atoms in macromolecules. Therefore, we define $R_{i,w} =$

TABLE 2: PDB[a] Codes of the 40 Proteins used to Determine $R_{i,p}$

mostly α motif	mostly β motif	mostly α/β motif	mostly α+β motif
1aca[b]	1cdb[b]	3trx[b]	1pba[b]
1eca	1hcc[b]	1ego[b]	2pna[b]
1cpc[c]	1atx[b]	1xis	1and[b]
351c	1bds[b]	3chy	1hev[b]
1osa	1egf[b]	2ctc	1gps[b]
4icb	2tgf[b]	1cse	1gat[b]
1rro	1tpm[b]	2trx[c]	2bop[c]
2ccy[c]	2rhe	1aba	1pgx
1mba	1noa	5p21	1ctf
1mbd	2sim	3dfr	2sn3

[a] Coordinate files from the Brookhaven National Laboratory Protein Data Bank (ref 81). The search was based on ref 82; only X-ray diffraction structures refined at high resolution (less than 1.7 Å) were considered; a total of 5610 residues and 51941 atoms were used in the search. [b] NMR structures. [c] All the polymer chains considered.

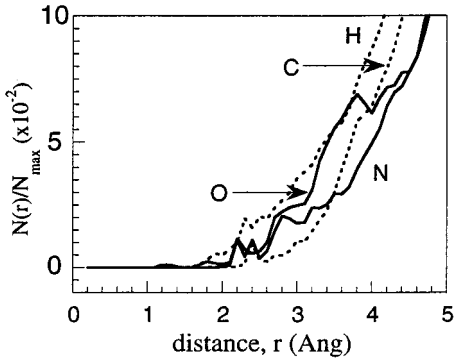

Figure 4. Radial distance distributions of atoms around atoms H and C (dashed lines), and N and O (solid line) in the sample used to define the cavity size (Born radius) within the protein medium. For clarity the distribution around sulfur atoms was omitted. The distribution was normalized to the maximum value N_{max} (see text).

$R_{i,COV} + h_i$ and $R_{i,p} = R_{i,COV} + g_i$ for the Born radii of atom i in the aqueous and protein phases, respectively. Here h_i and g_i are positive quantities accounting for the enlargement of the cavity defined by the covalent radii of the atoms in the two phases, respectively, and their values are given below. To be consistent with these definitions we take $R_{i,v} = R_{i,COV}$, that is, we assume that there is no extension of the covalent radius in vacuo. To assign values to h_i, in this preliminary evaluation, they were assumed to be independent of atom type, and the values of 0.35 Å and 0.85 Å were taken for negative and positive partial charges, respectively, as has been proposed previously.[54]

To determine the cavity size of an atom embedded within a protein the radial probability distributions were evaluated for each atom type from 40 structures taken from the Protein Data Bank.[81] The proteins were selected to be representative of several structural motifs[82] as indicated in Table 2 where the 40 proteins have been tabulated. Radial distribution functions $N(r)$ were calculated for each atom, where $r = d(X-Y)$ is the distance from an atom-type X = H, C, O, N, or S to any atom Y. Once the probability function, $P(r)$, for finding an atom Y at distance r from X is known, the mean value of r can be calculated from $\langle r \rangle = \Sigma r P(r)$, where the sum runs over the interval $0 < r < r_m$; r_m is defined by $N(r_m) = N_m$, where N_m is the value of N assumed to contain the first shell around the given atom X (see below). The probability function $P(r)$ is related to the radial distribution, $N(r)$, by $P(r) = ZN(r)/r^2$ where Z normalizes $P(r)$ in the interval $0 < r < r_m$. In Figure 4 the normalized distribution, $N(r)/N_{max}$, versus r is shown, where N_{max} is the maximum of the total distribution, and is located at about 16 Å for all atoms in the 40 proteins listed in Table 2. Only atoms belonging to the first-neighbor shell around a given atom are required, and it is sufficient to include the region below 5 Å. To obtain the value of g, the mean cavity radius around each atom must be estimated, and this was obtained by calculating the average distance, $\langle d \rangle$, from a given atom to all atoms with distribution between 2.5 and 5% of the maximum, N_{max}, of the total distribution, that is, $100(N_m/N_{max}) = 2.5$ (or 5). At 5% it was found that $\langle d \rangle$ is about 3 ± 0.6 Å, whereas at 2.5% $\langle d \rangle$ has a value of about 2.5 ± 0.6 Å, so that this range of the distribution approximately accounts for the first shell of neighbors around any atom in the protein. Note that the width of the distribution, σ, was calculated in the same way.

The average covalent radius is $\overline{R_{COV}} \sim 0.75$ Å[83] and the average of h, taking into account the sign of charges usually found in proteins, is ~0.7 Å. The value of g at the 2.5% level and at the 5% level is estimated from the relation $\langle d \rangle = 2\overline{R_{COV}} + \bar{h} + g$. The average value of g is around 0.5 Å for all atom types in the 40-structure database. Thus, in the present approximation, all atom types are enlarged by about the same extent when they are transferred from water into protein, that is, $g_i = h_i + g$. The standard deviations around the average values of d suggest that the use of a single value of g for all atoms is a reasonable first approximation for evaluating $R_{i,p}$. Nevertheless, the standard deviations are large enough to suggest that a useful refinement would be the resolution of $\langle d \rangle$ on the basis of atom type, and accounting for the different partial charges associated with each atom type similarly as has been done earlier for ions in water.[54] Further detailed studies of the Born radius showed that it depends on solvent discreteness and polarity; the relation between the Born radius and the solute−solvent distribution function was also analyzed.[79]

In the CHARMM force field hydrogen bonding is implicitly incorporated into the electrostatic component, as is also the case for most currently used force fields. However, in the SCP-ISM this implicit description of H-bonding is at least partially eliminated because with large damping ($D_s \sim 10$ at $r = 3$ Å) the electrostatic contribution will be substantially reduced. A second aspect of this problem is related to the evaluation of $\langle d \rangle$ from the distributions given in Figure 4. Implicit in the evaluation is that there is little or no interpenetration of van der Waals spheres, which is certainly true except for hydrogen bonding where the van der Waals sphere of the proton substantially penetrates the van der Waals sphere of the proton acceptor. This suggests that Born radii of protons participating in H-bonding may require a modified definition, and this is the approach we have followed to account for the overdamping of the H-bonding electrostatic interaction terms suggested above. An advantage of this approach is that the additional short- and medium-range interactions operative in H-bonding can be localized to their effective proton-acceptor separations. Moreover, geometric constraints can also be accounted for. The description of the hydrogen bonding algorithm that we have incorporated into the ISM is given in the immediately following paper.[84]

3. Parametrization of the SCP-ISM. To calculate the electrostatic energy of a molecule from eq 18, the parameters λ in the screening function $D_s(r)$ (cf. eq 10) must be first determined. In an earlier paper, an assignment of the screening parameter, α, where $\alpha = \lambda(\epsilon_s + 1)$, was given for each amino acid side-chain analogue that was based on the chemical functional groups of the amino acid side chains.[57] Here the evaluation of the screening parameters was completely determined in the context of the all-atom force field, PAR22,[58] in

CHARMM.[59] However, the procedure is general and can be applied in a straightforward way to any currently available force field. In PAR22, 35 different atom types have been defined to construct all the naturally occurring amino acids, and, therefore, a corresponding set of 35 screening parameters, α, had to be determined. An advantage of this approach is that the extension to other systems of biological relevance, for example, nucleic acids, is then straightforward. The parameters are determined by fitting experimental solvation free energies of amino acid side-chain analogues in water on the basis of the equation

$$\delta G = \sum_I (\Delta G_s - \Delta G_{s,exp})^2 \tag{23}$$

where $\Delta G_{s,exp}$ is the experimental solvation free energy of the molecules and ΔG_s is the calculated solvation free energy; the index I is understood to run over the 21 analogues considered for the parametrization (both neutral and charged His were included). The values of α are optimized to minimize δG. The Coulomb contribution to ΔG_s and E_T (right hand side of eqs 18 and 19) depends on a combined screening parameter α_{ij}. The screening parameters are calculated from the geometric means, that is, $\alpha_{ij} = \sqrt{\alpha_i \alpha_j}$ to ensure symmetry of the contributions to the interaction energy (eq 13), and in analogy with the form used for the treatment of the van der Waals parameters in the CHARMM force field.[59]

As previously discussed, classic dielectric theory and experimental evidence indicate that polar solvation sigmoidally screens electrostatic fields. Referring to Figure 1, it is evident that the slope of the sigmoid is a strong function of the bulk value of the dielectric constant of the solvent because the curves tend to reach their asymptotic values at the same value of r, regardless of the nature of the solvent. Thus for fluids with small bulk screening constants (3−5) the sigmoidal screening function will reach this value slowly [starting from $D(0) = 1$], suggesting they can be reasonably well-approximated by a constant value over the whole range of distances. Tests of this assumption (Hassan, unpublished results) shows that this is indeed the case, and therefore, the screening function in a vacuum D_v (state 2 in Figure 2), will be taken as constant.

In a preliminary determination of the α, D_v was set to 1 so that the polarization energy was calculated from the equation

$$\Delta G_s^{pol} = \frac{1}{2}\sum_{i\neq j}^{N} \frac{q_i q_j}{r_{ij}}\left[\frac{1}{D_s(r_{ij})} - 1\right] + \frac{1}{2}\sum_{i}^{N} q_i^2 \left\{\frac{1}{R_{i,Bs}}\left[\frac{1}{D_s(R_{i,Bs})} - 1\right]\right\} \tag{19'}$$

The fitting procedure yielded the parameter set given in column two of Table 3 and the regression between calculated and experimental values was $\Delta G_s = 0.97\Delta G_{s,exp} + 0.01$, with $R = 0.99$. Thus an excellent fit was achieved. The screening function D_s plotted in Figure 5 (labeled $D_v = 1$) is obtained from the average of α calculated from the values given in the second column of Table 3. This screening increases very slowly and even at 20 Å is still very far from the asymptotic limit. It clearly underscreens the Coulombic interaction given the radial permittivity function of water plotted in Figure 1, and the relation between $D_s(r)$ and $\epsilon(r)$ expressed by eq 11. Tests with this preliminary set of α values on several small peptides with conformational memories[60] simulations resulted in overly compact, rigid structures (Hassan and Konvicka, unpublished results).

The reciprocal of α is a measure of the average solvation radius of an atom,[57] and large values yield screening functions that reach their asymptotic values slowly, whereas small values give a compressed sigmoidal behavior that rapidly approaches its asymptotic value. The value of α^{-1} used to plot $D_s(r)$ (with $D_v = 1$ in Figure 5) is about 8 Å. However, recent pK_a calculations[72] indicate that α^{-1} should be about 1 Å for solvent-exposed groups. A reparametrization to achieve an $\langle\alpha^{-1}\rangle \approx 1$ Å necessitated relaxing the condition that $D_v = 1$, and a simulated annealing technique was used to find the minimum of δG[85] for the 35 α_i and 21 values of D_v (corresponding to the 20 naturally occurring amino acids with His in both charged and neutral form). The simulated annealing was performed by decreasing the values of the analogue of the temperature

TABLE 3: Atom-Type-Based Parameters α

atom type[a]	parameter α, $D_v = 1$	parameter α, $D_v > 1$[b]
CT1	0.0540	0.9618
CT2	0.0025	0.9618
CT3	0.0475	1.2314
C	0.0775	0.9802
CC	0.0925	1.0064
CA	0.0020	1.0461
CY	0.0584	0.8741
CPT	0.0584	1.1021
CPH1	0.2584	1.2524
CPH2	0.0917	0.9404
CP1	0.0540	1.1154
CP2	0.0540	1.0971
CP3	0.0540	1.1295
H	0.0098	1.2935
HA	0.0690	0.7822
HC	0.0590	0.8319
HP	0.0092	1.2227
HS	0.1097	0.8168
HR1	0.0750	1.1075
HR2	0.1760	1.1083
HR3	0.1770	1.0331
HB	0.0521	1.1197
NY	0.1430	0.7083
NC2	0.0370	0.7000
NH1	0.2173	1.2279
NH2	0.0317	0.8751
NH3	0.1500	1.2995
NR1	0.1250	1.1881
NR2	0.3584	1.1685
NR3	0.7700	1.2217
N	0.1370	0.9104
O	0.0222	0.9841
OC	0.4525	1.1637
OH1	0.0375	0.8004
S	0.0500	1.0219
average values	0.1192 ± 0.1475	1.0367 ± 0.1647

[a] PAR22 all-atom type definition (ref 58). [b] Both αs and D_vs optimized by simulated annealing MC (ref 84).

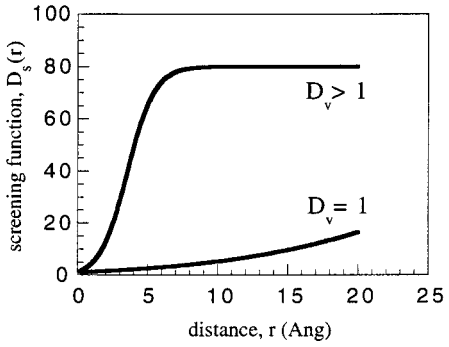

Figure 5. Screening function $D_s(r)$ as a function of the distance r obtained with the average values of α given in Table 3. The curve labeled $D_v = 1$ is obtained from the average values in column 2 of Table 3; the curve labeled $D_v > 1$ is obtained from the average value in column 3 of Table 3.

TABLE 4: Vacuum Screening Constants, D_v, for the 21 Amino Acid Analogues

amino acid	screening constant D_v	amino acid	screening constant D_v
Ala	7.11	Ser	3.84
Ile	6.84	Gln	3.65
Leu	7.94	Asn	3.78
Val	7.43	Thr	4.13
Asp^-	1.72	Cys	3.02
Glu^-	1.67	Met	3.15
His^+	1.87	Phe	3.66
Lys^+	1.00	Tyr	3.46
Arg^+	2.70	His	3.31
Gly	1.89	Trp	3.33
Pro	4.63		

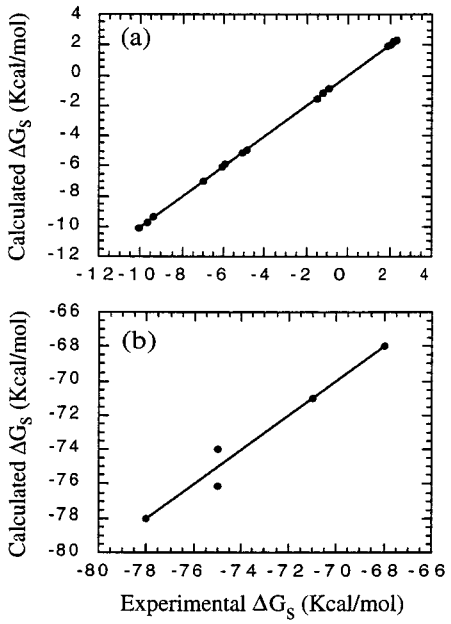

Figure 6. Scatter plot of experimental and calculated solvation free energies of the neutral (a) and charged (b) amino acid side-chain analogues, obtained with the MC simulated annealing optimization (ref 85) of both α and D_v (third column in Table 3 and Table 4, respectively).

parameter τ in the weighting factor $\exp(-\delta G/\tau)$ of the Boltzmann distribution. The sampling distribution of every α was restricted to $0.6 < \alpha < 1.3$ to ensure that the restraint $\langle\alpha^{-1}\rangle \approx 1$ Å was achieved. The resulting screening parameter values are given in the third column of Table 3 and the D_vs are given in Table 4. The D_s with this $\langle\alpha^{-1}\rangle = 0.96$A is also plotted in Figure 5. The sigmoidal form of this curve clearly lies in the range of the radial dielectric permittivity of water, as shown in Figure 1. The values of D_v shown in Table 4 are in good agreement with the interior dielectric permittivity of analogues of amino acids that have been reported previously.[86]

Figure 6A and B give the scatter plots between the experimental and calculated solvation energies of the neutral and charged analogues, respectively. As shown above from the initial determination of the α, and as also reported by other authors,[29] a good fit to solvation energies does not guarantee a reliable implicit solvent model for structural studies. To simultaneously reproduce experimental solvation energies, and reproduce the theoretically correct screening (see Figure 1), the atomic screening parameters had to be restricted to a narrow range. The fit to the experimental solvation energies is excellent, as seen from Figure 6. As shown in the following paper,[84] the present formulation of an implicit solvent model was able to calculate structures of small peptides, given the sequence only, that were in good agreement with experiment. It is also important to note that each reformulation of the model, for example, describing it with a different number of parameters, PAR19[87] for example, will require careful testing and analysis to assess if the model still provides a reliable representation of bulk solvent effects for its intended applications.

III. Comparison of Energetics from SCP-ISM with PB

To obtain an initial indication of the reliability of the SCP-ISM, a comparison of various energetic quantities has been carried out with those obtained from solution of the PB equation. Such comparisons among different theoretical continuum models can be of considerable help to detect deficiencies and virtues of the different techniques. In this work the FDPB numerical calculations have been solved using the University of Houston Brownian dynamics (UHBD) package kindly provided by Dr. J. A. McCammon.[88,89] Results are presented from calculations of the polar solvation free energy and its components for MC-generated conformations of several peptides ranging from a single amino acid to a nonapeptide. To be consistent, the FDPB calculations have also been carried out using the PAR22 parameters even though the calculated polar solvation energies for amino acid analogues were somewhat less accurate than with some of the other parameter sets.[86,90] Therefore the ΔG_s calculated with the ISM (for which the solvation energies of the amino acid analogues are essentially the experimental values, see Figure 6) are, in general, shifted from the values calculated with the FDPB. Despite this, the correlation between the calculated solvation energies from different conformations of the peptides obtained from the two techniques provides a measure of the consistency of the two approaches. Similar comparisons between the GB/SA (Surface Area) and the FDPB were reported recently.[91] The FDPB calculations were carried out with the UHBD program using an internal dielectric constant, ϵ_i, of 1 and a solvent dielectric of 80; the ionic strength was set to 0, the temperature was 300 K, and a probe radius of 1.4 Å was used. The system was divided into a 65 × 65 × 65-grid box with cell length ranging from 0.2 to 0.65 Å according to the size of the solute.

1. Self- and Interaction Energies. Because of the fundamental difference in the way the self-energies are calculated in the SCP-ISM and FDPB approaches, it is of interest to consider separately the correlation of the two contributions to the solvation energy, that is, ΔG_s^{elec} and ΔG_s^{self}. For this purpose the dipeptide Tyr-Glu^- was constructed in CHARMM, with standard C- and N-termini, and 40 different conformations were generated by a MC search at high temperature. To simplify the calculation and to have control over the individual interactions involved, the partial charges of all the particles in the system were set to 0, except the charges at the coordinates of three atoms. These three atoms were chosen so that their interparticle distances ranged between 2 and 9 Å in the set of 40 conformations, sampling all of the critical regions of $D_s(r)$. The charges were placed on the Tyr backbone oxygen ($q = 1$), on one of the Glu carboxyl oxygens ($q = -1$), and on one of the hydrogen atoms bound to the C_β of Tyr ($q = -1$); therefore, the net charge on the peptide was -1.

The individual terms of ΔG_s^{pol} with the UHBD program were obtained in the following way: to estimate the self-energy contribution of each atom in its particular position in every structure, the polar solvation energy of the whole dipeptide was computed with the other two charges set to 0. The total ΔG_s^{self} of the system was then obtained as the sum of the three individual self-energies. Next, the total ΔG_s^{pol} of the peptide,

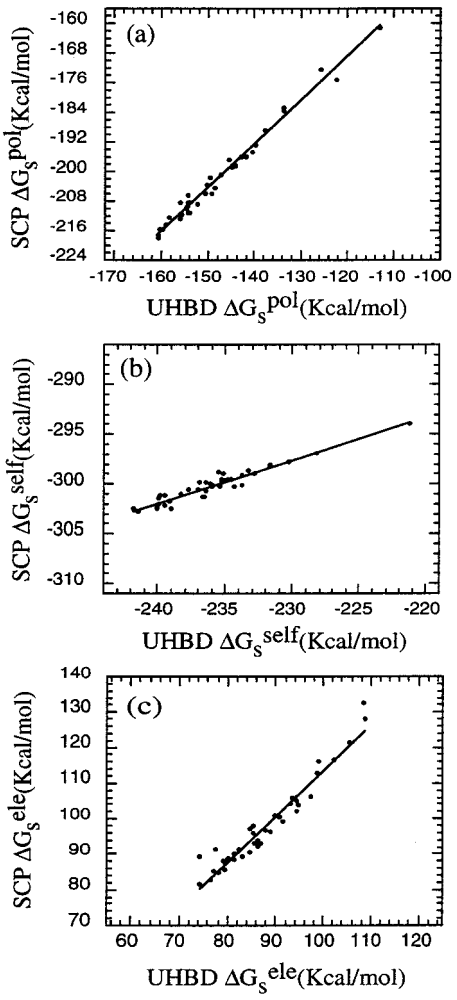

Figure 7. Scatter plot of 40 different conformations of YE of the polar component of the solvation energy (a), ΔG_s^{pol}; the self-energy term (b), ΔG_s^{self}; and the electrostatic interaction term (c), ΔG_s^{elec}, between the SCP-ISM and UHBD (refs 88, 89) PB calculation.

that is, with the three charges switched on, was calculated for every conformation, and the total electrostatic interaction ΔG_s^{elec} was obtained from the difference $\Delta G_s^{pol} - \Delta G_s^{self}$.

The contributions from the SCP-ISM were calculated from eq 24. For all the results discussed below, the values of the screening constant $D_v(r_{ij})$, appearing in the interaction term of eq 24, were obtained as the geometric means of the individual amino acid values, that is, $D_v = (D_{vi}D_{vj})^{1/2}$. This simple assumption is not intended to be the optimal approach for calculating the vacuum screening in any system, but it is consistent with the treatment of the parameters α in the screening function D_s. In fact it was observed that both the correlation coefficient as well as the slope of the fitting line is sensitive to the way D_v for the coupling terms is defined. It is clear that further analysis must be carried out to obtain a physically reasonable way to assign D_v (r_{ij}) in any arbitrary macromolecule once the individual D_vs of the composing amino acid residues are known.

Figure 7 shows the scatter plots of these three quantities, that is, ΔG_s^{pol} (Figure 7A), ΔG_s^{self} (Figure 7B), and ΔG_s^{elec} (Figure 7C). Figure 7B and C show that the individual contributions to the polar solvation energy are highly correlated (0.96 and 0.97 for the self- and electrostatic energies, respectively). This remarkable result is particularly striking for ΔG_s^{self} because this quantity is calculated in a completely different way in the SCP-ISM from that in the FDPB. For the FDPB, the changes in self-energy in the different conformations is solely due to the changes in molecular shape defined by the way the interior and solvent permittivities are assigned. For the SCP-ISM, the changes in self-energy in the different conformations depend on the changes in the degree of exposure of each atom in the different conformations, which are reflected in the effective Born radii. It is also noted that the high correlation observed for ΔG_s^{elec} confirms the essential reasonableness of the way $D_v(r_{ij})$ has been defined, although better choices could undoubtedly be found. These results demonstrate *the validity of the way the effective Born radius has been defined and the fundamental validity of the relation between the SCP and LDS theory*. It is also of interest that the correlation of ΔG_s^{pol} is larger ($R = 0.99$) than the correlations of the individual contributions, ΔG_s^{elec} and ΔG_s^{self}. This is due to the way the model was parametrized by fitting total solvation free energies. It is important to appreciate, however, that the high correlation of ΔG_s^{pol} would not require the two contributing components to be highly correlated, because even if they had shown low correlation, they could have compensated each other in such a way to yield a highly correlated ΔG_s^{pol}.

TABLE 5: Peptide Sequences, Net Charge, and Statistics from Scatter Plots between SCP-ISM and FDPB Polar Solvation Energies

peptide	termini, Q^a	standard deviation[b]	maximum deviation[b,c]	r^d
K	zw, 1	1.42	7.1 (10.1)	0.92
YE	n, −1	0.49	2.6 (2.6)	0.93
YEA	n, −1	0.69	3.5 (3.2)	0.96
YEADG	n, −2	2.46	11.1 (4.8)	0.89
ANYEWHGVL	n, 0	5.54	31.5 (14.0)	0.89
ENYEWHGEL	n, −2	4.07	21.4 (6.5)	0.95
EEEEEEEEE	n, −9	4.00	21.5 (1.5)	0.93

[a] zw, zwitterion with standard amino and carboxyl termini; n, neutral termini; Q, net charge on peptide. [b] kcal/mol. [c] Numbers in parentheses are maximum distributions, in percent of the mean value, around the regression line. [d] Correlation coefficient.

2. Polarization Free Energies. The solvation energies of seven peptides, ranging from a single amino acid to three nonapeptides, and different net charges have been calculated with the SCP-ISM and the FDPB using UHBD.[89] The sequences, correlations, and standard deviations from the scatter plots are given in Table 5. The scatter plots of ΔG_s^{pol} calculated with the SCP-ISM and FDPB are presented in Figure 8 for Lys, the dipeptide, tripeptide, and pentapeptide, and for the nonapeptide the scatter plots are given in Figure 9. The correlation of ΔG_s^{pol} calculated from the two methods is 0.9 or better for all the peptides studied here. Moreover, there seems to be no systematic variation in the correlations depending on either peptide length or net charge. The scatter plots show that the maximum width of the distribution around the regression line lies between 2 and 15% (see Table 5), but also does not seem to depend on peptide length or charge. The standard deviations listed in Table 5 lie within the experimental errors of the measured hydration free energy of the amino acid analogues.[76]

Comparison of the scatter plots in Figure 8A−D emphasizes the remarkable independence of the scatter and correlation to the size of the peptide. This is also the case for the nonapeptides presented in Figure 9A−C. At the same time it is clear from Figures 8 and 9 that the slopes of the regression lines are not 1, nor are the intercepts 0. These effects tend to increase with increasing net charge and increasing size of the peptide, as has also been observed in similar comparisons with the GB/SA approximation and the FDPB.[91] These shifts have no consequence for structure calculations where the relevant quantities

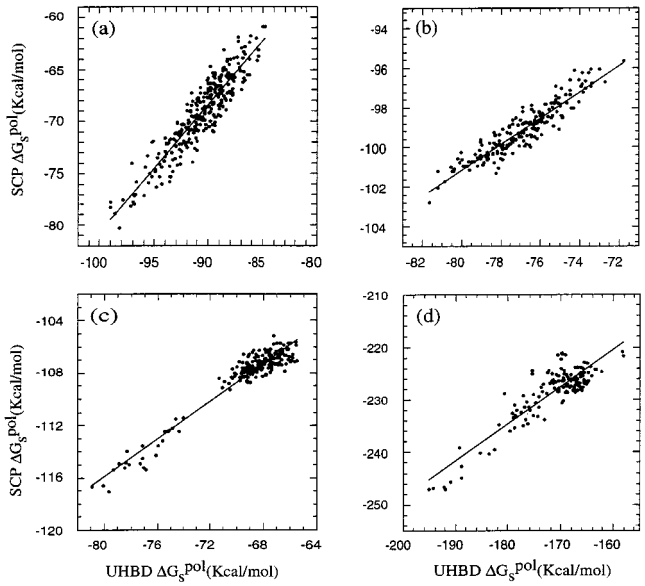

Figure 8. Scatter plots of the polar solvation free energy ΔG_s^{pol} between the SCP-ISM and the UHBD PB calculation; (a) 300 structures of single amino acid Lys^+ (zwitterionic, net charge +1); (b) 200 structures of the dipeptide YE (neutral ends, net charge −1); (c) 200 structures of the tripeptide YEA (neutral ends, net charge −1); (d) 140 structures of the pentapeptide YEADG (neutral ends, net charge −2).

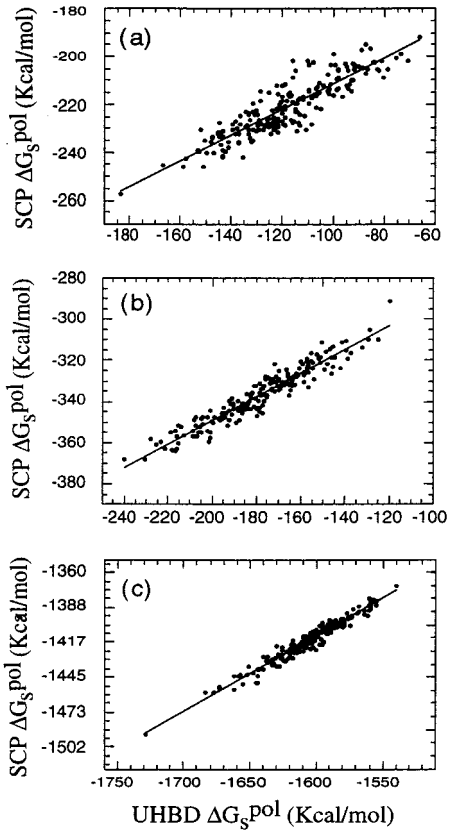

Figure 9. Scatter plots for three different nanopeptides (200 conformations each); (a) ANYEWHGVL (neutral termini, net charge 0); (b) ENYEWHGEL (neutral termini, net charge −2); (c) EEEEEEEEE (neutral ends, net charge −9).

are differences in total energy, but it is certainly of interest to understand their origin. From Figure 7 it is seen that the main source of these differences is due to the self-energy component. This is not surprising, as this component is the most difficult to calculate properly by any method. In contrast, the difference in the interaction term, ΔG_s^{elec}, is much less, moreover, the slope

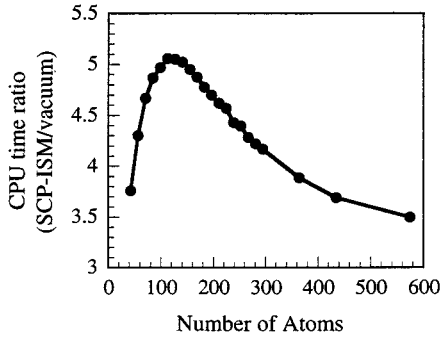

Figure 10. Ratio of CPU time between the SCP-ISM and the vacuum calculations in CHARMM as a function of the number of atoms *N* in the system.

of this regression line is about 1.3. Thus the offset in the total polarization energy (see Figure 7A) is primarily due to the self-energy, but the slope of the regression is about 1.2.

IV. CPU Time

To obtain some initial insight into the CPU requirements for the SCP-ISM, timing data for a dodecapeptide[92,93] is reported and compared with results from MD calculations (Sankararamakrishnan, unpublished results) with CHARMM.[59] All calculations were carried out on a Silicon Graphics platform using a single R10000 chip, with 5000 energy evaluations carried out for each case. The times are given relative to the vacuum calculation, which required 90 s. ISM calculations without computation of the SASA increased CPU time by a factor of 1.6. Incorporating the analytic SASA calculation in CHARMM increased the time by an additional factor of 3.2. Therefore we incorporated into CHARMM an optimized version of the SASA calculation used in GEPOL,[94,95] which required an additional factor of about 2.5. Thus the complete ISM calculation is only about four-fold longer than the vacuum calculation. Immersing the dodecapeptide into a water sphere of 30 Å radius (required to ensure that all atoms of the dodecapeptide are fully immersed using the default cutoff radius in PAR22[58]) added about 3800 water molecules to the system. The calculation of 5000 energies required about 40 times more CPU time than the full ISM calculation so that even for this small system very significant savings in CPU time are achieved. A more extensive analysis is presented in Figure 10. Vacuum and SCP-ISM calculations have been carried out for polypeptides up to 50 amino acid residues with infinite cutoff radius. The plot in Figure 10 shows the ratio of CPU time between the SCP-ISM and the vacuum case. This ratio increases to a maximum value of about 5 at around 120 atoms and then decreases with increasing system size to a factor of about 3. The fact that the ratio approaches a limiting value shows that the dependence of the CPU time on size is the same in the SCP-ISM as in the vacuum.

Conclusions

A new formulation of an implicit solvent model has been proposed that is based on screened Coulomb potentials. The screening function is nonlinear, distance dependent, and of sigmoidal form, which is the correct form derived from basic electrostatics as formulated in the LDS theory of polar solvation and from experiment. Reaction field corrections are easily incorporated into the formulation. To obtain the correct dielectric behavior it was necessary to introduce the notion of internal screening (the D_v) with values greater than unity. Interestingly the values found from the MC optimization procedure were between 2 and 7, in good agreement with the values found for

this parameter using other models.[96−98] The appearance of the internal screening in eq 19 follows from carrying out the thermodynamic cycle exhibited in Figure 2 without a priori assumptions. It will not appear if it is assumed that $D_v \equiv G_i = 1$.

To test consistency of the approach with other models the polar component of the solvation energy was calculated for several peptides with differing net charge as a function of conformation and compared with results from solving the numerical PB equation. The results were found to be highly correlated. A decomposition of the polar contribution showed that the interaction and self-energy components, separately, were also highly correlated. This latter result is particularly significant for the self-energy term because it is calculated in entirely different ways in the SCP-ISM and the FDPB equation.

Ultimately, the validity of the SCP-ISM must be established by showing that experimental results can be predicted. Although the fundamental formalism presented in this paper is an important step in attaining this goal, it is not sufficient. We show in the accompanying paper that when an interpenetration model of hydrogen bonding is also incorporated, then sequence-to-structure simulations can be achieved.

Acknowledgment. We thank Dr. Karel Konvicka for help in incorporating conformational memories into CHARMM. Computational support was provided by the Pittsburgh Supercomputer Center (sponsored by the National Science Foundation), the Cornell National Supercomputer Facility, the Advanced Scientific Computing Laboratory at the Frederick Cancer Research Facility of the National Cancer Institute (Laboratory for Mathematical Biology), and the University Computer Center of the City University of New York. Partial support of the work by NSF grant DBI 9732684 and NIH grant DA-09083 (S.A.H.) are gratefully acknowledged.

References and Notes

(1) Luecke, H.; Schobert, B.; Richter, H.-T.; Cartailler, J.-P.; Lanyi, J. K. *Science* **1999**, *286*, 255−260.
(2) Bryant, R. G. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 29−53.
(3) Ooi, T. *Adv. Biophys.* **1994**, *30*, 105−154.
(4) Ben-Naim, A. *Hydrophobic Interactions*; Plenum Press: New York, 1980.
(5) Teeter, M. M. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 577−600.
(6) Teeter, M. M. *Dev. Biol. Stand.* **1992**, *74*, 63−72.
(7) Jensen, L. H. *Dev. Biol. Stand.* **1992**, *74*, 53−61.
(8) Doniach, S.; Eastman, P. *Curr. Opin. Struct. Biol.* **1999**, *9*, 157−163.
(9) Mehler, E. L. The Lorentz−Debye−Sack Theory and Dielectric Screening of Electrostatic Effects in Proteins and Nucleic Acids. In *Molecular Electrostatic Potential: Concepts and Applications*; Murray, J. S., Sen, K., Eds.; Elsevier Science: Amsterdam, 1996; Vol. 3, pp 371−405.
(10) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283−422.
(11) Warshel, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211−217.
(12) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1988**, *9*, 327−335.
(13) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301−332.
(14) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591−3600.
(15) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415−436.
(16) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219−10225.
(17) Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101−1109.
(18) Tazaki, K.; Doi, J. *J. Phys. Chem.* **1996**, *100*, 14520−14525.
(19) Hoijtink, G. J.; de Boer, E.; van der Meer, P. H.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas* **1956**, *75*, 487−503.
(20) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(21) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213−217.
(22) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem.* **1997**, *101*, 3005−3014.
(23) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(24) Hendsch, Z. S.; Sindelar, C. V.; Tidor, B. *J. Phys. Chem.* **1998**, *102*, 4404−4410.
(25) Liang, J.; Subramaniam, S. *Biophys. J.* **1997**, *73*, 1830−1841.
(26) Resat, H.; Marrone, T. J.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 522−532.
(27) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(28) Tironi, I. G.; Luty, B. A.; van Gunsteren, W. F. *J. Chem. Phys.* **1997**, *106*, 6068−6075.
(29) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem.* **1997**, *101*, 8098−8106.
(30) Friedrichs, M.; Zhou, R.; Edinger, S. R.; Friesner, R. A. *J. Phys. Chem.* **1999**, *103*, 3057−3061.
(31) Warshel, A.; Russell, S. T.; Churg, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 4785−4789.
(32) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem.* **1997**, *101*, 4458−4472.
(33) Warshel, A.; Aqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267−298.
(34) Born, M. *Z. Phys.* **1920**, *1*, 45−48.
(35) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305−8311.
(36) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8552−8554.
(37) Schaefer, M.; Bartels, C.; Karplus, M. *J. Mol. Biol.* **1998**, *284*, 835−848.
(38) Collura, V. P.; Greaney, P. J.; Robson, B. *Protein Eng.* **1994**, *7*, 221−233.
(39) Given, J. A.; Gilson, M. K. *Proteins* **1998**, *33*, 475−495.
(40) Luo, R.; Head, M. S.; Moult, J.; Gilson, M. K. *J. Am. Chem. Soc.* **1998**, *120*, 6138−6146.
(41) Dominy, B. N.; Brooks, I. C. L. *J. Phys. Chem.* **1999**, *103*, 3765−3773.
(42) Webb, T. J. *J. Am. Chem. Soc.* **1926**, *48*, 2589−2603.
(43) Debye, P.; Pauling, L. *J. Am. Chem. Soc.* **1925**, *47*, 2129−2134.
(44) Schwarzenbach, G. *Z. Phys. Chem. A* **1936**, *176*, 133−153.
(45) Harvey, S. C.; Hoekstra, P. *J. Phys. Chem.* **1972**, *76*, 2987−2994.
(46) Pennock, B. D.; Schwan, H. P. *J. Phys. Chem.* **1969**, *73*, 2600−2610.
(47) Takashima, S.; Schwan, H. P. *J. Phys. Chem.* **1965**, *69*, 4176−4182.
(48) Hasted, J. B.; Ritson, D. M.; Collie, C. H. *J. Chem. Phys.* **1948**, *16*, 1−21.
(49) Debye, P. *Polar Molecules*; Dover: New York, 1929.
(50) Lorentz, H. A. *Theory of Electrons*; Dover: New York, 1952.
(51) Sack, V. H. *Phys. Z.* **1926**, *27*, 206−208.
(52) Sack, V. H. *Phys. Z.* **1927**, *28*, 199−210.
(53) Ehrenson, S. *J. Comput. Chem.* **1989**, *10*, 77−93.
(54) Bucher, M.; Porter, T. L. *J. Phys. Chem.* **1986**, *90*, 3406−3411.
(55) Mehler, E. L.; Eichele, E. *Biochemistry* **1984**, *23*, 3887−3891.
(56) Mehler, E. L. *J. Phys. Chem.* **1996**, *100*, 16006−16018.
(57) Guarnieri, F.; Schmidt, A. B.; Mehler, E. L. *Int. J. Quantum Chem.* **1998**, *69*, 57−64.
(58) MacKerell, A. D. J.; Bashford, D.; Bellot, M.; Dunbrack, R. L. J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Wiorkiewicz-Kuczera, J.; Karplus, M. *Biophys. J.* **1992**, *6*, A143.
(59) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(60) Guarnieri, F.; Wilson, S. R. *J. Comput. Chem.* **1995**, *16*, 648−653.
(61) Guarnieri, F.; Weinstein, H. *J. Am. Chem. Soc.* **1996**, *118*, 5580−5589.
(62) Toulouse, M.; Fritsch, V.; Westhof, E. *Mol. Simul.* **1992**, *9*, 193−200.
(63) Nayfeh, M. H.; Brussel, M. K. *Electricity and Magnetism*; John Wiley & Sons: New York, 1985.
(64) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486−1493.
(65) Böttcher, C. J. F. *Physica* **1938**, *V*, 635−639.
(66) Pollock, E. L.; Alder, B. J.; Pratt, L. R. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 49−51.
(67) King, G.; Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1991**, *91*, 3647−3661.
(68) Bjerrum, N. *Z. Phys. Chem. Stoechiom. Verwandschaftsl.* **1923**, *106*, 219−241.
(69) Kirkwood, J. G.; Westheimer, F. H. *J. Chem. Phys.* **1938**, *6*, 506−512.

(70) Conway, B. E.; Bockris, J. O. M.; Ammar, I. A. *Trans. Faraday Soc.* **1951**, *47*, 756−767.
(71) Jackson, J. D. *Classical Electrodynamics*; John Wiley & Sons: New York, 1975.
(72) Mehler, E. L.; Guarnieri, F. *Biophys. J.* **1999**, *77*, 3−22.
(73) Warshel, A.; Aqvist, J.; Creighton, S. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5820−5824.
(74) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem.* **1997**, *101*, 5239−5248.
(75) Schmidt, A. B. *Biophys. Chem.* **1994**, *51*, 393.
(76) Schmidt, A. B.; Fine, R. M. *Mol. Simul.* **1994**, *13*, 347−365.
(77) Tannor, J. D.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A.; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875−11882.
(78) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588−5593.
(79) Papazyan, A.; Warshel, A. *J. Phys. Chem.* **1998**, *102*, 5348−5357.
(80) Latimer, W. M.; Rodebush, W. H. *J. Am. Chem. Soc.* **1920**, *42*, 1419−1433.
(81) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rogers, J. R.; Kennard, D.; Simanouchi, T.; Tatsumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.
(82) Przytycka, T.; Aurora, R.; Rose, G. D. *Nat. Struct. Biol.* **1999**, *6*, 672−682.
(83) Pauling, L. *The Nature of the Chemical Bond*; Cornell University Press: Ithaca, 1960.
(84) Hassan, S. A.; Guarnieri, F.; Mehler, E. L. *J. Phys. Chem. B* **2000**, *104*, 6490.
(85) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. *Science* **1983**, *220*, 671−680.
(86) Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1990**, *94*, 7684−7692.
(87) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(88) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187−197.
(89) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewics, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. *Comput. Phys. Commun.* **1995**, *91*, 57−95.
(90) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(91) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *J. Phys. Chem.* **1997**, *101*, 1190−1197.
(92) Hassan, S. A.; Guarnieri, F.; Mehler, E. *Biophys. J.* **1999**, *76*, A198.
(93) Hassan, S. A.; Guarnieri, F.; Mehler, E. *Biophys. J.* **2000**, *78*, 335A.
(94) Pascual-Ahuir, J. L.; Silla, E. *J. Comput. Chem.* **1990**, *11*, 1047−1060.
(95) Pascual-Ahuir, J. L.; Silla, E.; Tunon, I. *J. Comput. Chem.* **1994**, *15*, 1127−1138.
(96) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1999**, *103*, 6142−6156.
(97) Pethig, R. *Dielectric and Electronic Properties of Biological Materials*; Wiley & Sons: Chichester, 1980.
(98) Sham, Y. Y.; Muegge, I.; Warshel, A. *Biophys. J.* **1998**, *74*, 1744−1753.