PROTEINS: Structure, Function, and Genetics 4:7–18 (1988)

Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis

Michael K. Gilson and Barry Honig
Department of Biochemistry and Molecular Biophysics, Columbia University, New York, 10032

ABSTRACT **In this report we describe an accurate numerical method for calculating the total electrostatic energy of molecules of arbitrary shape and charge distribution, accounting for both Coulombic and solvent polarization terms. In addition to the solvation energies of individual molecules, the method can be used to calculate the electrostatic energy associated with conformational changes in proteins as well as changes in solvation energy that accompany the binding of charged substrates. The validity of the method is examined by calculating the hydration energies of acetate, methyl ammonium, ammonium, and methanol. The method is then used to study the relationship between the depth of a charge within a protein and its interaction with the solvent. Calculations of the relative electrostatic energies of crystal and misfolded conformations of *Themiste dyscritum* hemerythrin and the VL domain of an antibody are also presented. The results indicate that electrostatic charge-solvent interactions strongly favor the crystal structures. More generally, it is found that charge-solvent interactions, which are frequently neglected in protein structure analysis, can make large contributions to the total energy of a macromolecular system.**

Key words: protein electrostatics, conformational energy, solvation

INTRODUCTION

The total electrostatic energy of a macromolecular system can be partitioned into contributions from interactions between pairs of charged atoms on the macromolecule and from the interactions of individual charges with the solvent.[1] Charge-solvent interactions are important factors in the energetics of protein folding and in the binding of charged substrates to macromolecules. For example, the strong favorable interactions between ionized amino acids and the aqueous solvent account for the tendency of these groups to be found on or near the protein surface.[2,3] In the past few years a number of studies have demonstrated that finite difference solutions to the Poisson-Boltzmann (PB) equation (the FDPB method) provide a reliable means of obtaining charge-charge interactions in macromolecules.[4–9] In this study we show how the FDPB method can be extended to the calculation of charge-solvent interactions as well.

The simplest and best-studied example of electrostatic charge-solvent interactions is the hydration of monatomic ions. The Born model[10,11] has been used for many years as a means of calculating enthalpies and free energies of hydration and has recently been shown to be quite accurate when consistently applied.[12] Although its original derivation was based on the nonintuitive physical process of charging and discharging an ion, as shown below and discussed previously,[13] the Born expression can be derived purely from reaction field considerations that have a clear physical basis. That is, the electrostatic contribution to the solvation free energy of an ion corresponds exactly to the interaction of the ion with the polarization it induces in the solvent. Unfortunately, the simplicity of the Born expression is lost for ions lacking spherical symmetry. However, this paper describes how the FDPB method can be used to calculate the interaction of a charge with its induced polarization for molecules of complex charge distribution and shape. This makes it possible, with only moderate computational cost, to obtain electrostatic solvation energies of molecules, electrostatic solvation contributions to the energies of free and bound substrates, and the total electrostatic energy of a macromolecule, explicitly accounting for solvation/desolvation effects.

Our implementation of the FDPB method has been discussed in a number of recent publications.[6,8,9,14] The protein is described in terms of its three-dimensional structure with all atoms located at positions defined by the crystal structure or generated by model-building techniques. In order to account, in an average way, for the effects of electronic polarizability, atomic charges are considered to be embedded in a low dielectric medium. The surrounding solvent is assigned an appropriate dielectric constant, most frequently taken to be that of water, and may contain an electrolyte obeying the Poisson-Boltzmann equation [see, e.g., ref. 10].

Received March 14,1988; revision accepted May 11, 1988.

Address reprint requests to Dr. Barry Honig, Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168 St., New York, NY 10032.

As mentioned above, the FDPB method has provided a useful basis for the interpretation of experimental results on charge-charge interactions in a number of complex macromolecular systems. The method has not previously been applied to the calculation of charge-solvent interactions, in part because there has been no way to extract ion solvation energies from the electrostatic potential maps produced by the FDPB method. Rashin and Namboodiri [13] have recently used a boundary element technique to solve the Poisson equation for small polar molecules and have thus avoided the problem of extracting solvation energies from finite difference solutions. However, given the relative ease of application of the FDPB method to large molecules such as proteins and nucleic acids and the importance of accounting for ionic strength effects, which are not easily included in boundary element methods, we have preferred to extend the applicability of the FDPB method to the calculation of solvation effects.

The present work describes the method that has been developed for this purpose. We begin by reviewing the definition of the total electrostatic energy of a molecule in classical electrostatics, emphasizing the separation into various charge-charge and charge-solvent interaction terms. The computational method is then described, and its accuracy is tested against analytic solutions to the linearized Poisson-Boltzmann equation.

The new methodology is applied first to the calculation of hydration energies of several small molecules. Good agreement with experimental hydration free energies is obtained using standard parameter sets. However, it should be possible to improve the accuracy of the results with some effort in developing parameters appropriate for solvation. Two applications to protein energetics are also presented: 1) an exploration of charge-solvent interaction energies in two proteins, and of their dependence upon the depth of an atom from the protein's surface; and 2) calculations of the charge-solvent interaction energy and total electrostatic energy of correctly and incorrectly folded conformations of two proteins. These studies are intended to demonstrate the wide applicability of the method and to illustrate the magnitude of charge-solvent interactions that are often neglected in the analysis of conformational and binding energies.

MATERIALS AND METHODS

Theory

Model

In a previous work[1] we defined the total electrostatic energy of a protein in terms of the interactions of atoms which were treated as charged spheres. This approach made it possible to include solvation ("self-energy") terms in an expression for the total electrostatic energy which was based on continuum electrostatics. However, in this work, instead of using shells of charge, we use the standard molecular mechanics description of charged atoms as point charges centered at the nucleus. Each atom is treated as a low-dielectric sphere whose dielectric constant depends on the electronic polarizability of the atom. It should be emphasized that since with this description no charge is ever closer than a van der Waals radius from the surface of an atom, and hence from a dielectric boundary, discontinuities that arise in classical electrostatics from point charges at boundaries are never encountered. For any configuration of atoms, the total electrostatic energy becomes that of a set of point charges, q_i, each embedded in a low-dielectric cavity corresponding to its atomic radius. Taken together, the atomic radii define the low-dielectric cavity of the macromolecule, which is assigned a uniform dielectric constant ϵ_m. The atoms are surrounded by a solvent of dielectric constant ϵ_s, which may contain an ionic atmosphere.

Terms contributing to the total electrostatic energy

It is convenient to partition the total electrostatic energy into a contribution from the Coulombic interaction of the charges with each other, ΔG^o_c, a contribution caused by the interaction of the charges with a polarizable solvent, ΔG^o_p, and a contribution caused by the interaction of the charges with the ion atmosphere in the solvent, ΔG^o_a [see reference 1]. Because the ion atmosphere introduces special complications if the nonlinear PB equation is used, in this work we will limit ourselves to the linearized form. The nonlinear PB equation will be considered in a future study.

For simplicity of notation we combine the solvent polarization and ion atmosphere terms into a single charge-solvent interaction term, $\Delta G^o_s = \Delta G^o_p + \Delta G^o_a$. The total electrostatic energy can now be written as:

$$\Delta G^o = \Delta G^o_c + \Delta G^o_s = \sum_i (\Delta G^o_{c,i} + \Delta G^o_{s,i}) \quad (1)$$

where the subscript i denotes the contribution of each charge. For the linearized PB equation it is straightforward to show that:

$$\Delta G^o = 1/2 \sum_i q_i (\phi_{c,i} + \phi_{s,i}) \quad (2)$$

where $\phi_{c,i}$ and $\phi_{s,i}$ denote, respectively, the Coulombic potential and the potential produced by the solvent at the location of charge i. The precise meaning of these terms can best be understood once an appropriate reference state has been introduced.

Reference State

It proves to be particularly convenient to define a reference state in which all charges are infinitely separated in a medium corresponding to the dielectric constant of the molecule, ϵ_m, and containing no ion atmosphere. All energies and potentials in this work are defined relative to this reference state. The ad-

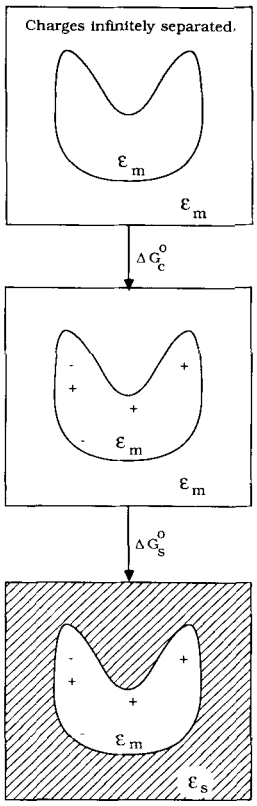

Fig. 1. Thermodynamic process for calculation of a molecule's total electrostatic energy.

vantage of using this state is that the Coulombic term, ΔG_c^o, is zero since all charges are infinitely separated, while the solvent term, ΔG_s^o, is zero because there are no dielectric boundaries in a homogeneous medium and the ionic strength has been set to zero.

The reference state is shown in the top panel of Figure 1. Notice that the surface of the molecule is drawn in the figure but that this does not correspond to a dielectric boundary since the interior of the molecule and the surrounding medium have the same dielectric constant, ϵ_m. Figure 1 shows a pathway for assembling the charges from the reference state into their final configuration in the solvated macromolecule. The individual steps in this process are discussed in the following sections.

The Coulombic term

In the first step of the process, the charges are brought together in the medium of dielectric constant ϵ_m. Note that the molecular surface still does not correspond to a dielectric boundary. Since this medium is homogeneous, Coulomb's law is valid. The free energy change in this step, ΔG_c^o is thus given by the Coulombic expression:

$$\Delta G_c^o = \sum_i \Delta G_{c,i}^o = 1/2 \sum_i \sum_{j \neq i} q_i q_j / \epsilon_m r_{ij} \qquad (3)$$

The solvent term

After the charges have been assembled the solvent environment is formed. ΔG_s^o, the charge-solvent interaction energy of the system, corresponds to the work performed in bringing the solvent boundary from infinity to the position defined by the surface of the molecule. The total electrostatic free energy change of the system, ΔG^o, is then given by equation 1. The central contribution of this paper is a new method for calculating ΔG_s^o for complex systems.

It is instructive to decompose ΔG_s^o into its individual contributions:

$$\Delta G_s^o = \sum_i (\Delta G_{s,ii}^o + \sum_{j \neq i} \Delta G_{s,ij}^o) \qquad (4)$$

where the first term describes the interaction of charge i with its own induced polarization (reaction field), while the second describes the interaction of charge j with the polarization induced by charge i.

The first term in equation 4 can be thought of as the sum of the solvation energies of each of the individual charged atoms in the configuration in which they are embedded in the macromolecule. Each element, $\Delta G_{s,ii}^o$, in the sum corresponds to the charge-solvent interaction energy of a particular charged atom assuming all other atoms to be neutral. It should be emphasized, however, that the other atoms still contribute to the overall shape of the macromolecule and hence to the solvent accessibility of atom i.

The second term in equation 4 accounts for solvent screening of the Coulombic interaction between charges i and j, as previously discussed.[1] Solvent screening is frequently incorporated into an effective dielectric constant, for example the distance dependent dielectric constant used in molecular mechanics calculations. A comparison of the distance-dependent dielectric constant with the predictions of the present model has been presented.[9]

Special case of a single ion —the Born model

The derivation of an expression for the electrostatic solvation energy of a single spherical ion serves as a useful illustration of the theory developed above. Consider, for example, the transfer of an ion from vacuum ($\epsilon = 1$) to water ($\epsilon = 80$). For simplicity we asume zero ionic strength. To obtain the transfer energy, we first calculate the electrostatic energy of the ion in water and in vacuum, both relative to the reference state. The transfer energy is then just the difference between these two values.

The two energies can be obtained from the process depicted in Figure 1. The reference state consists of a single point charge immersed in a continuous medium of dielectric constant ϵ_m. The molecular surface in this case is defined by a sphere whose radius is that of the ion. The Coulombic term, ΔG_c^o, is zero since the system contains only one charge. Similarly, there is no ij term (see equation 4) in ΔG_s^o. Thus, the total electrostatic energy, ΔG^o, will be given by $\Delta G_{s,ii}^o$.

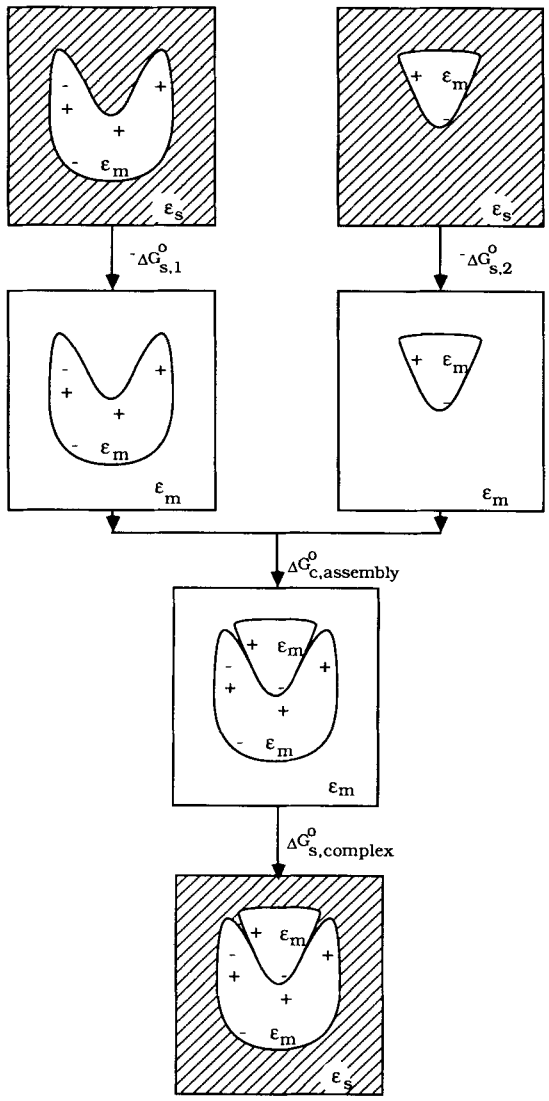

Fig. 2. Themodynamic process for calculation of change in total electrostatic energy upon intermolecular binding.

As shown in equation 2, $\triangle G^o$ is determined by the electrostatic potential at the center of the ion. In keeping with the present formalism, this potential for a single ion may be termed $\phi_{s,ii}$. From classical electrostatic theory, the potential at charge, q_i, located at the center of a sphere of dielectric constant ϵ_m, which is embedded in a medium of dielectric constant ϵ_s, is given by:

$$\phi_{s,ii} = (332\ q_i/a_i)\ (1/\epsilon_s - 1/\epsilon_m) \qquad (5)$$

In equation 5, a_i is the ion's radius, and the constant 332 produces a result in kcal/mol. The electrostatic energy is thus given by:

$$\Delta G^o = (166\ q_i^2/a_i)\ (1/\epsilon_s - 1/\epsilon_m) \qquad (6)$$

Substituting $\epsilon_s = 80$ for water and $\epsilon_s = 1$ for vacuum and taking the difference between the two resulting values for the ion's total electrostatic energy yields a transfer energy of $(166\ q_i^2/a_i)\ (1/80 - 1)$, which will be recognized as the Born expression.[10,11] Note that this expression has been obtained without invoking nonphysical charging and discharging processes. Rather, the electrostatic transfer energy results entirely from differences in charge-solvent interactions, which correspond to the second step in Figure 1. It is worth pointing out that in this example, the transfer energy is independent of ϵ_m, but that this will be the case only for spherically symmetric boundaries and charge distribution.

It should be pointed out that in the more general instance where the object being transferred is not a single spherical ion but rather is an entire molecule with a complex shape and charge distribution, the potentials cannot be obtained from equation 5 and must be calculated numerically (see below).

Conformational energies

The reference state used here is well defined for a given polypeptide sequence. The electrostatic energy of the unfolded state is poorly defined but is irrelevant when the energies of different folded conformations are being compared. As a consequence, the total electrostatic energies of different conformations of a given polypeptide may be compared in order to determine their relative electrostatic stabilities. However, the electrostatic energies of proteins having different sequences are not comparable in this way, because the reference states and the unfolded states of two different polypeptides are different.

Binding energies

The change in total electrostatic energy when two molecules form a complex may be calculated by means of a three-step thermodynamic process (Fig. 2) similar to that used in calculating the total energy of a single molecule. The initial state has both molecules fully solvated and infinitely separated from each other. In the first step, the solvent around each molecule is replaced by a medium having the dielectric constant of the molecular interior (ϵ_m) and zero ionic strength. The work involved is $(\Delta G^o_{s,1} + \Delta G^o_{s,2})$, where 1 and 2 refer to the two separate molecules. Coulomb's law, with $\epsilon = \epsilon_m$, may now be used to compute the work of assembling the two molecules to form the complex (ΔG^o_s assembly). Finally, the solvent is added back to yield the fully solvated complex. The work of performing this step is $\Delta G^o_{s,complex}$.

An alternate description of binding that may offer better insight in some circumstances is illustrated in Figure 3, which shows a two-step thermodynamic process whose initial and final states are the same as those in the previous process. In step 1, each molecule is partially desolvated by removing the solvent from the region that the other molecule will come to occupy

in the complex, and replacing the removed solvent by the usual medium of dielectric constant ϵ_m and zero ionic strength. The work involved ($\Delta\Delta G^o_{s,1} + \Delta\Delta G^o_{s,2}$) is the change in the charge solvent interaction energies of the two species upon binding. Thus, if species 2 is a calcium ion, $\Delta\Delta G^o_{s,2}$ corresponds to the loss of solvation energy of the calcium upon binding to the protein. In the second step, the charges of molecule 1 are transferred to the low-dielectric space prepared next to molecule 2 in the previous step. The work in this step is the interaction energy between the charges of the two molecules, calculated with the complex surrounded by solvent (ΔG^o_{inter}). (Actually, for this process to be properly defined the charged and uncharged species are switched in the middle panels of Fig. 3. There is of course no electrostatic contribution from the uncharged species).

This second description of the electrostatics of binding has the advantage of yielding numbers that are readily interpretable, since the binding energy is separated into the change in the hydration energy of each molecule plus the solvent-screened interaction energy of the two molecules.

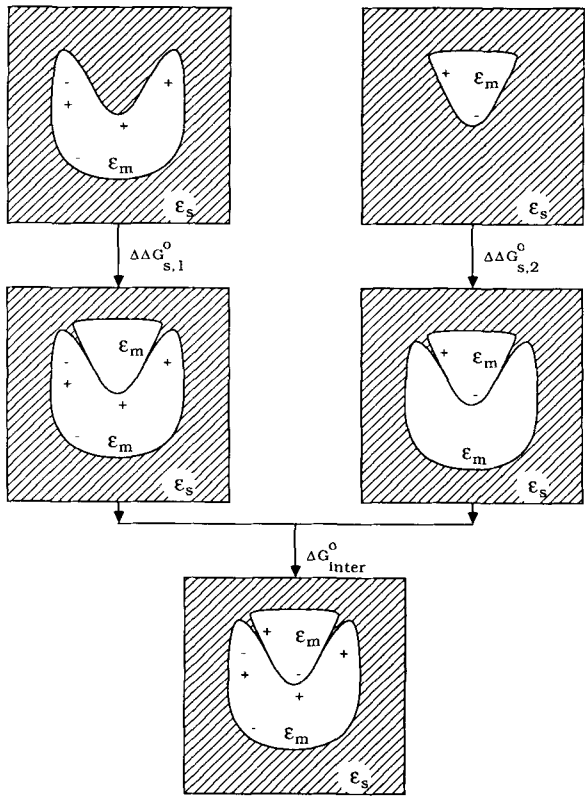

Fig. 3. Alternative thermodynamic process for calculation of change in total electrostatic energy upon intermolecular binding.

Computational Method

The computational development that makes the present work possible is a procedure for calculating the change in solvent-interaction of a molecule caused by an alteration in the dielectric constant of the solvent. For an arbitrary change in the solvent, let the initial solvent dielectric constant be ϵ_1 and the final value be equal to ϵ_2. Two finite difference calculations[6,14] must be performed. In both, the charges are at their correct locations in the protein, and in both, the dielectric constant of the protein region is set to its normal value, ϵ_m. However, in one calculation the region surrounding the protein is defined by ϵ_1, and in the second it is defined by ϵ_2. The difference in electrostatic potential at each charged atom for these two calculations must result entirely from the change in solvent parameters. It should be pointed out here that the finite difference method yields large, seemingly arbitrary potentials at the locations of charges. However, the differences in these potentials are meaningful and correspond to the change in the solvent polarization energy of the system upon altering the environment. This is given by:

$$\Delta G^o_s = 1/2\Sigma_i q_i \Delta\phi_i \qquad (7)$$

where $\Delta\phi_i$ is the change in calculated potential at charge q_i upon altering the solvent. It should be noted that the calculations can be performed in a completely analogous fashion to obtain changes in electrostatic energy when the ionic atmosphere is changed.

In the applications described in the "Theory" section, where one is interested in determining the total electrostatic energy of a protein in a given conformation and solvent relative to the reference state, ϵ_1 must be equal to ϵ_m. Then equation 7 yields the work involved in the second step of the thermodynamic process described in Figure 1. The energy of the first step, assembly of all the charges in a medium of ϵ_m, is given by straightforward application of Coulomb's law as discussed above and is carried out analytically.

The finite difference calculations described here were carried out using the software package DelPhi[9,14] which solves the Poisson-Boltzmann equation. Two techniques may be used to reduce the errors that result from replacing continuous functions by functions on the finite difference grid. In the "focusing" technique [14,15] two or more finite difference calculations are performed, in which each calculation uses the same number of grid vertices, but the grid is made finer with each run by reducing the spatial extent of the lattice. The boundary conditions for each calculation (except the first) are obtained from the results of the preceding coarser calculation, and the series of runs thus "focuses in" on the molecule under study. This approach provides a relatively fine grid, together with accurate boundary conditions. The second technique, rotational averaging,[14] involves per-

forming a series of finite difference calculations, where each calculation has the physical system under study rotated by a different angle with respect to the finite difference grid. The electrostatic potential fields from the runs are then averaged to yield results of improved accuracy. The focusing and rotational averaging techniques may be combined with each other.[14]

Since it has been shown that accurate solutions to the linearized Poisson-Boltzmann equation can be obtained for charges as close as ~1 Å to a dielectric boundary or from another charge,[14] it appears likely that the potential difference at a charge produced by altering the solvent can be accurately calculated if the solvent remains at least 1 Å from the charge. This condition should be easy to meet, as the solvent cannot approach closer than the van der Waals radius of an atom, and the smallest van der Waals radius, that of hydrogen, is about 1 Å.

Atomic and Molecular Parameters

A set of atomic radii and a solvent probe radius are required in order to define the dielectric boundary between a molecule and the solvent. As discussed below, it would be desirable to develop a parameter set that reproduces the experimentally observed solvation energies of biologically important ionic species. However, since such a parameter set is not yet available, in this work standard van der Waals radii are used.

The study of charge-solvent interactions as a function of charge depth below the surface is based on the same set of "united atom" van der Waals radii[16] which were used in recent studies on subtilisin.[8,9] The other calculations in this paper treat polar hydrogens bound to nitrogens and oxygens explicitly. These hydrogens are given a radius of 1.2 Å[17] and the nitrogens and oxygens are given radii of 1.5 Å[17] and 1.6 Å,[16] respectively.

Unless otherwise specified, the polar hydrogen parameter set 19 from the molecular mechanics program CHARMM[18] provided the atomic charges used in this work. The CHARMM charges used for methyl ammonium, acetate, and methanol were obtained from the amino acid side chains lysine, aspartate (or glutamate), and serine, respectively. The charge set for ammonium was derived from that for the lysine side chain by transferring 0.1 proton charges from the nitrogen to the hydrogen replacing the methylene group, thus producing a symmetrical charge distribution.

Protein Structures

The Brookhaven Protein Data Bank [19] provided the atomic coordinates of rhodanese (1RHD),[20] crambin (1CRN),[21] *Themiste dyscritum* met-hemerythrin (chain A of 1HMQ),[22] and the VL domain of an Fab immunoglobulin fragment (residues 1–113 of 1MCP).[23] The energy-minimized and misfolded structures of hemerythrin and VL domain, which were very generously provided to us by Drs. J. Novotny and B. Bruccoleri, have been fully described and analyzed by a number of techniques [24] and are described briefly in "Results and Discussion" below. The misfolded conformations used here were generated by a torsional search over side chain conformations followed by energy minimization, using the program CONGEN.[25]

A problem that arose in the conformational analysis of the hemerythrin and VL domain structures was the placement of hydrogen atoms, since the coordinates of hydrogens are not determined by X-ray crystallography. (Hydrogen bonding is treated in CHARMM and CONGEN as at least partly electrostatic.) On the other hand, hydrogen coordinates were included in some of the computationally generated structures that were provided to us. In order to place all of the structures on an equal footing in this regard, all were stripped of any hydrogens, and polar hydrogens were added back using the CHARMM command HBUILD to generate a first set of structures. A second set of structures was obtained by energy-minimizing the first set of structures with only the hydrogens free to move. The minimizations were performed in 200 conjugate gradient steps using CHARMM's default energy parameters, and updating the nonbonded lists every ten steps. The RMS changes in all hydrogen positions resulting from minimization were roughly 0.2 Å. The two sets of structures are discussed further in "Results and Discussion."

The electrostatic calculations on hemerythrin and VL domain assumed the chain termini (including the artifical C-terminus at residue 113 of the VL domain) to be charged, all histidines to be neutral, and all aspartates, glutamates, lysines, and arginines to be fully ionized. In accord with the work of Novotny and coworkers, [24,26] the iron atoms in hemerythrin have been omitted from all calculations. No distance cutoff was used in applying Coulomb's law. The Coulombic interactions reported in this study exclude all interactions between atoms connected by one or two covalent bonds (1–2 and 1–3 interactions), but include all interactions between atoms connected by three or more bonds (1–4 interactions, etc.). The purpose of the exclusions is to avoid large artifactual changes in interaction energies associated with slight variations in bond lengths and angles, which should be accounted for completely by bonded energy terms. 1–4 interactions were included because distances between atoms joined by three covalent bonds may change significantly with dihedral angle. For example, the distance between two main chain nitrogen atoms changes from 2.9 Å to 3.8 Å when the intervening ψ angle varies from 0° to 180°.

TABLE I. Accuracy Tests*

b	q	Foc	Rot	Final grid size (Å)	Numerical results (kcal/mol)	Tanford-Kirkwood results (kcal/mol)	Error (%)
30	(1,29,0,0)	N	Y	1.36	−36.0	−40.6	13
30	(1,12.9,24.9,7.5)	N	Y	1.36	−38.0	−40.6	5
30	(1,12.9,24.9,7.5)	Y	Y	1.00	−42.5	−40.6	4
30	(1,29,0,0)	Y	Y	1.00	−38.9	−40.6	4
4	(1,0,0,0) (1,0,2.5) (1,0,−2.5) (−1,2.5,0,0) (−1,−2.5,0,0)	N	N	0.18	−56.9	−60.0	5

*Comparisons between charge-solvent interaction energies calculated using Tanford-Kirkwood theory and using the numerical method described in this paper. b, Radius of low-dielectric sphere centered at origin; q, charges, where first value is charge in atomic units, and remaining three values are Cartesian coordinates. Foc, Y if two-step focusing used, N if not; Rot, Y if sixfold rotational averaging used, N if not; Final grid size, distance between vertices of finite difference grid. (If focusing used, this value corresponds to the most detailed grid.) See text for further details.

RESULTS AND DISCUSSION

Accuracy Tests

The numerical technique for calculating charge-solvent interaction energies was tested by comparisons with the results of analytic solutions for a low-dielectric sphere surrounded by a high-dielectric solvent containing an electrolyte.[1,27] A number of cases were tested. In each case, the sphere's internal dielectric constant was set to 2; the final external dielectric constant and ionic strength were 80 and 0.15 M, respectively, while the initial external dielectric constant and ionic strength were 2 and 0.0 M and correspond to the state illustrated in the second panel of Figure 1. The thickness of the Stern layer [10,14] was set to zero.

Table I presents the results of these tests. The first four have a single charge 1 Å below the surface of a 30 Å sphere, while the last test has five charges arranged in a 4 Å sphere, forming a small "ion." As shown in Table I, it is possible to calculate charge-solvent interaction energies accurate to within about 5% for each of these systems. Note that the single-charge cases have charges very close to the sphere's surface, taxing the accuracy of the numerical method.[14] The results for deeper charges are equivalent or better (results not shown).

Based on the results for the model "ion," the calculations of hydration energies of small molecules, presented below, were performed using no focusing and no rotational averaging (see "Materials and Methods" for descriptions of these techniques) but a small grid size. The other charge-solvent calculations presented in this paper also used no focusing, but did use rotational averaging over six angles. Although focusing would have improved the accuracy somewhat, this improvement did not appear to be worth the additional computer time given the qualitative emphasis of this work.

Comparison With Experiment: Hydration Energies of Small Molecules

As a means of testing the physical model, we have used the method described above to calculate the electrostatic component of the charge-solvent interaction energies of four polar compounds whose hydration energies are known from experiment: methyl ammonium, ammonium, acetate, and methanol.

Two finite difference calculations were performed for each molecule: one with an external dielectric constant of 80 (water) and one with an external dielectric constant of 1 (vacuum). The internal dielectric constants were taken to be 2. In order to speed convergence, physiological ionic strength was used together with the external dielectric constant of 80. This appears to have a negligible effect on the calculated hydration energies, since redoing one of the acetate calculations using zero ionic strength changed the hydration energy by only 0.1%. As a further test, we repeated the continuum hydration enthalpy calculations of Rashin and Namboodiri[13] for acetate, obtaining results that agree within 3%.

The calculated electrostatic free energies of hydration for methyl ammonium, ammonium, acetate, and methanol are presented in Table II, along with experimentally determined hydration free energies. The comparison between calculated and experimental values is complicated by the fact that there are no direct experimental measurements of individual ions. Rather, the experimental values that are reported depend upon the assumptions used to obtain the hydration free energy of a single reference ion, ΔG_s°(ion), where "ion" refers to either H^+ or K^+ (see refs. 28,29).

TABLE II. Calculated Electrostatic Component of Hydration Free Energies, Compared With Experimental Hydration Free Energies, for Three Small Molecules*

	Hydration free energy (kcal/mol)
Methyl ammonium	
Experimental**	−71. (−63.)
Calculated	−77.
Ammonium	
Experimental**	−79. (−71.)
Calculated	−86.
Acetate	
Experimental†	−79. (−87.)
Calculated	−70.
Methanol	
Experimental‡	−5.1
Calculated	−4.3

*All values correspond to hypothetical 1 M reference states in vapor and aqueous phases. For the ions, the first values correspond to a proton hydration free energy ($\Delta G_s^\circ(H^+)$ of −262.5 kcal/mol[28,29,36–38] and the parenthesized values correspond to $\Delta G_s^\circ(H^+ = -254.3)$ kcal/mol, suggested by Marcus.[29] The gas phase data used in calculating the ionic hydration energies have uncertainties of about ±3 kcal/mol.
**Ref. 39.
†Calculated using thermodynamic cycle described and referenced by Scheraga and coworkers[33] except that the hydration free energy of neutral acetic acid was obtained from Ref. 40.
‡Ref. 41.

These single ion hydration energies are subject to considerable uncertainty and tend to vary by approximately 10 kcal/mol. Table II emphasizes the experimental results associated with $\Delta G_s^\circ(H^+) = -262.5$ kcal/mol, which appears to be the most generally accepted value (see footnote to Table II for references) and which yields better agreement with our calculations. However, the experimental results obtained for $\Delta G_s^\circ(H^+) = -254.3$ kcal/mol[29] are also included.

The calculated values differ from the experimental values calculated using $\Delta G_s^\circ(H^+) = -262.5$ kcal/mol by 8–19%. (With $\Delta G_s^\circ(H^+) = -254.3$ kcal/mol, the maximum error rises to 24%, for acetate) These results are encouraging, especially in light of the fact that the atomic parameters — standard van der Waals radii, and the CHARMM charge set -- have not been optimized for the present method of calculating hydration free energies. On the other hand, the calculations yield only the electrostatic contribution to the energies, so a comparison with experimental results is not entirely valid. Dispersion forces and the hydrophobic effect, for example, are likely to make contributions of the same order of magnitude as the differences between the calculated and experimental values.

As previously noted by Rashin and Namboodiri [13] in a study of hydration enthalpies, the continuum electrostatic model appears capable of yielding results whose quality equals that of more computationally demanding molecular simulations. The present results are similar to those obtained in free energy simulations of amino acids in water.[30] Direct comparison of results is not possible, however, since somewhat different molecules were studied. Ultimately, it should be possible to improve the electrostatic free energies provided by the continuum model by finding atomic charges and radii that reproduce the experimental values for hydration free energies of amino acid side chains, metal ions, and other groups of biological interest. However, it appears that even without further paramaterization, the FDPB approach introduced in this work may be expected to yield reasonable estimates of solvent polarization energies. In the following sections we use the method to determine the magnitude of these terms in a number of different systems.

Applications to Proteins

Charge-solvent interactions as a function of charge depth

Although burying a charged atom within a protein incurs a substantial penalty in solvation energy, a buried charge may still interact significantly with the solvent. Conversely, an atom whose surface is accessible to solvent may yet be partially desolvated. In this section the magnitude of charge-solvent interaction energies are calculated for a number of atoms, and the relationship between these energies and the distances between the charges and the solvent is examined.

The solvent interaction energies of 45 charged atoms in the large protein rhodanese (293 residues), and ten charged atoms in the small protein crambin (46 residues), were calculated under the assumption that all other atoms in the protein were neutral. The results therefore yield the interaction of each atom with its own induced polarization in the solvent. In order to facilitate comparison, the atoms examined were considered to bear unit charges. (Their solvent-interaction energies may be corrected for a nonunit charge q by multiplying the values provided here by q^2.) Each atom's depth from the surface was also calculated by determining its distance from the closest point of a Connolly dot surface [31] generated using the same parameters as used in defining the protein-solvent dielectric boundary (see "Materials and Methods").

The energies were calculated using an internal dielectric constant of 2, an initial external dielectric constant of 2, an initial ionic strength of 0.0 M, a final external dielectric constant of 80, and a final external ionic strength of 0.15. Thus the energies are determined relative to that of a charge in the reference

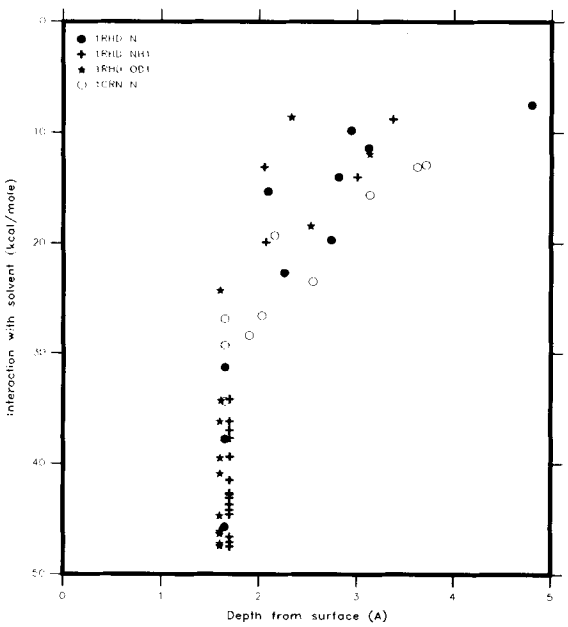

Fig. 4. Charge-solvent interaction energies as a function of atomic depth, for proteins rhodanese (1RHD) and crambin (1CRN). N, main chain nitrogen; NH1,Nη1 of arginine; OD1,0δ1 of aspartate.

state, which here is assumed to have $\epsilon_m = 2$. These energies may also be regarded as electrostatic solvation energies of the charges as they are found in the protein, relative to the solvation energies they would have if they were completely buried in an infinitely large protein. Repeating the present calculations with an initial ionic strength of 0.0 M has very little effect ($\sim$1%) on the energies. No Stern layer was used, so mobile ions in the solvent approach to the protein's dielectric boundary.

The selected atoms for rhodanese are the main chain nitrogens of residues 1, 29, 39, 54, 86, 111, 113, 116, 129, 249; all 14 aspartate OD1 atoms; all 20 arginine NH1 atoms, and CD1 of isoleucine 270. The isoleucine atom was selected because it is the deepest (7.6 Å) atom in the protein when a 1.4 Å probe is used to generate the surface. The crambin atoms examined are the main chain nitrogens of residues 1, 7, 13, 17, 23, 27, 31, 35, 40, and 46.

As shown in Figure 4, the solvent-interaction energies fall between 0 and -50 kcal/mol. The less negative values correspond to atoms that are more deeply buried and so are further from the solvent. The fact that atoms at the surface can have solvent-interaction energies ranging from ~ -50 kcal/mol to ~ -25 kcal/mol (Fig. 4) results from the present definition of depth, which does not distinquish between atoms almost completely surrounded by solvent, and atoms with only a small solvent-exposed patch.

The atoms with the strongest solvent interactions tend to be those belonging to the ionizable side chains—i.e., OD1 of Asp and NH1 of Arg— because these atoms are usually well solvated. The main chain nitrogen with a very negative energy belongs to the N-terminus, which is exposed to solvent; and the OD1 and NH1 atoms with the weakest interactions with solvent (~ -9 kcal/mol) form an unusual buried salt bridge.[20] The results for atoms not in direct contact with the solvent should be insensitive to their assumed atomic radii because these atoms do not contribute to the van der Waals surface of the protein. On the other hand, the fact that an atom is not in direct contact with the solvent does not mean that its interaction with the solvent is negligible. For example, the main nitrogen of residue 113 in rhodanese is 4.8 Å below the surface, but has an interaction with the solvent of -7.5 kcal/mol; and CD1 of isoleucine 270, 7.6 Å deep, retains -6.7 kcal/mol of interaction with the solvent. These results suggest that in the sulfate-binding protein of *S. typhimurium*, , whose structure was solved by Pflugrath and Quiocho[32] the sulfate, although buried, retains a substantial interaction with the solvent.

Figure 4 also demonstrates that the solvent interaction energies for crambin tend to be more favorable than those for rhodanese for a given charge depth. This must result from crambin's small size, which causes each atom to be nearer the solvent than an atom of similar nominal depth in a larger protein.

The fact that buried charges may retain substantial electrostatic interactions with the solvent suggests that desolvation costs will tend to be overestimated by free energy functions[33,34] which determine the work of dehydrating a charged atom based on measures of the atom's accessibility to solvent. A method of treating electrostatic charge-solvent interactions that would account for their long range and yet be computationally simple might serve as a useful correction to such free energy functions.

Conformational energy minimization, and misfolded proteins

The total conformational energy of a protein is known to be the sum of many large contributions such as hydrophobic interactions, chain entropy, electrostatics, etc. In attempting to distinguish between correctly and incorrectly folded polypeptide conformations it is important to have as accurate a treatment as possible for each of these terms. The methodology introduced in this work provides a realistic treatment of electrostatic interactions and should thus be useful in conformational analysis. In fact, Moult and James[35] have found that the electrostatic energy alone can be used to select stable conformations of loops in a trypsin-like enzyme from *Streptomyces griseus*.

They accounted for solvent screening of pairwise interactions using image charges (thus approximating the second term in equation 4), but did not account for the solvation of individual charged atoms

TABLE III. Electrostatic Energies of Hemerythrin and VL Domain for Crystal Structures, Energy-Minimized Structures, and Misfolded Structures*

	Solvent		Coulombic		Total	
	Absolute	Relative	Absolute	Relative	Absolute	Relative
Hemerythrin						
Crystal	−764.	0.	−741.	0.	−1,505.	0.
Minimized	−691.	+73.	−918.	−177.	−1,609.	−104.
Misfolded	−618.	+146.	−873.	−132.	−1,491.	+14.
VL Domain						
Crystal	−569.	0.	−566.	0.	−1,135.	0.
Minimized	−512.	+57.	−688.	−122.	−1,200.	−65.
Misfolded	−496.	+73.	−773.	−207.	−1,269.	−134.

*Energies in kcal/mol.

(the first term in equation 4). In this section we illustrate the application of the FDPB method to the calculation of electrostatic contributions to conformational energies of entire proteins. We also address the question of whether relative electrostatic energies can be used to identify stable protein conformations.

The proteins studied are hemerythrin and the VL domain of the immunoglobulin Fab fragment MC/PC 603 (see "Materials and Methods"). For each protein, the conformations analysed were[24] the crystal structure, a conformation generated from the crystal structure by energy minimization using CHARMM, and a severely misfolded conformation, generated using the program CONGEN. As described by Novotny et al.,[24] the misfolded conformation of hemerythrin was produced by assigning to its main chain atoms the coordinates of the main chain atoms of corresponding residues of the VL domain and then using the iterative torsional search program CONGEN[25] to optimize the side chain conformations. This procedure is made possible by the fact that hemerythrin and the VL domain have the same number of residues (113). The misfolded form of the VL domain was produced[24] by the converse process of giving its main chain the hemerythrin fold and optimizing its side chains in this new structure.

We used the methods described above to calculate each structure's total electrostatic energy, which consists of a solvent-interaction term and a Coulombic term. The calculations use an interior dielectric constant of 2, a solvent dielectric constant of 80, a solvent ionic strength of .15 M (physiological), and no Stern layer. As noted in "Materials and Methods," all 1–2 and 1–3 interactions were omitted in calculating the Coulombic terms.

Table III presents the electrostatic energies calculated for each conformation of the two proteins. The first column contains the interactions with solvent, relative to the interactions with a medium of dielectric constant 2; the second column contains the Coulombic energies of assembling the charges from the reference state. The third column contains the sums of the first two columns and represents total electrostatic energies relative to the reference state defined in "Materials and Methods." The absolute energies are given, as are the energies relative to the corresponding values for the crystal structure of each protein. The entries are averages of results from two sets of structures of each protein, one set with hydrogens added using the CHARMM command HBUILD, and the other with the hydrogens subsequently energy-minimized (see "Materials and Methods"). Although these two sets of structures yield electrostatic energies that differ by up to tens of kcal/mol, the signs and rankings of the entries in Table III do not depend on whether the hydrogens have been energy-minimized.

The calculations are subject to a number of errors. The charge-solvent interaction energies could be in error by on the order of 20%, based on the comparisons between calculated and experimentally determined hydration energies of small molecules, presented above. Further errors may result from the assumptions made concerning the ionization states of titratable residues, particularly in the case of histidines, which have been assumed neutral, and of which there are seven in hemerythrin, and one in the VL domain; from the neglect of hemerythrin's irons; and, as discussed below, from the treatment of hydrogen bonding.

In considering the results in Table III, we presume that for each protein, the most "correct" structure is the crystal structure; the next most correct is the energy-minimized crystal structure; and the misfolded structures are completely incorrect. It is striking that for both proteins, the solvent interaction energies rank the structures in order of correctness: the crystal structures show the strongest interaction with solvent; the energy-minimized structures have solvent-interactions of intermediate strength; and the misfolded structures have the least favorable interactions with solvent. It is interesting in this regard that Novotny et al.[24] found that misfolded hemerythrin had nine buried ionizable groups, while misfolded VL domain had only four. The significant loss

of solvent polarization energy in misfolded hemerythrin is likely to be due to the large number of buried ionizable groups.

In contrast to the solvent interaction energies, the Coulombic energies cannot be simply related to the "correctness" of a particular structure. Indeed for the VL domain, they are inversely correlated with the expected order of correctness. As a consequence, the total electrostatic energies are also ranked incorrectly. This would appear to suggest, in contrast to the conclusion of Moult and James,[35] that electrostatic energies alone cannot be used to identify stable protein conformations. However, the following considerations indicate that Table III should not be interpreted in this way.

First, the α-helical conformations (crystal and energy-minimized hemerythrin and misfolded VL domain) contain more hydrogen bonds than the β-sheet conformations (misfolded hemerythrin and crystal and energy-minimized VL domain). Since the Coulombic energies were calculated for all charge-bearing atoms including those participating in hydrogen bonds, it is not surprising that the α-helical conformations have the lowest Coulombic energies. However, the electrostatic component of hydrogen bonds is not well defined, and the use of partial charges to obtain hydrogen bonding potentials is, to a great extent, a matter of convenience. For this reason, it would be of interest to evaluate the total electrostatic energy in the absence of hydrogen bonding contributions.

A second problem is that it is not really meaningful to compare the energies of the crystal and energy minimized conformations, even if they contain identical secondary structures. Energy minimizations tend to produce overly compact structures since they neglect conformational entropy. These structures will have excessively negative Coulomb energies because, for example, hydrogen bond energies will be optimized by the minimization procedure. Thus, although the total electrostatic energies of the minimized structures are lower than those of the presumably more correct crystal structures, total electrostatic energies may still be useful in comparing the stabilities of different computer-generated conformations, all of which, for example, might be energy minimized.

Given the various complications, our results do not at this stage allow us to draw conclusions concerning the suitability of the total electrostatic energy as a criterion for identifying the relative stability of different protein conformations. However, our calculations do suggest that solvent interaction energies can help identify unstable conformations. This result, taken together with the loop conformation study of Moult and James,[35] suggest that total electrostatic energy may provide a useful measure of stability, if the complications associated with different types of secondary structure can be resolved. On the other hand, it should be emphasized here that there is no a priori reason to use electrostatic energy as a measure of stability since the total free energy of a protein is the sum of many contributions (see the report by Novotny el al.[24] for a discussion of other factors relating to hemerythrin and the VL domain). In any case, the methods introduced in this work now make it possible to account in a consistent fashion for both solvent screening of pairwise interactions and solvation energies in the analysis of protein conformations.

CONCLUSIONS

In this paper we have shown that the FDPB method can be extended to account for the effects of electrostatic charge-solvent interactions. In previous work we have shown how the method can be used to calculate solvent-screened charge-charge interactions. Thus, it is now possible to calculate the total electrostatic energy of a given macromolecular system. This work describes how the methodology can be applied to the calculation of the free energies of substrate binding and protein folding. Other possible applications include the study of molecular recognition and the improvement of energy functions used in molecular mechanics.

The numerical results presented in this work are intended to demonstrate the magnitude of various solvent polarization effects in a qualitative sense, and to illustrate how the FDPB method can be used in different applications. However, if truly quantitative agreement with experiment is to be obtained it will be necessary to develop a set of atomic parameters, which will yield appropriate hydration energies for compounds representative of the chemical groups found in proteins, such as carboxylates, ammoniums, and alcohols. In the meantime, the method makes it possible to obtain reasonable estimates of the rather large solvent polarization terms that are frequently neglected in conformational analysis molecular mechanics simulations.

ACKNOWLEDGMENTS

We are very grateful to Dr. J. Novotny and Dr. R. Bruccoleri for providing us with the energy-minimized and mis-folded protein structures. We also thank Drs. R. Bruccoleri, R. Fine, H. Rodman Gilson, J. Novotny, A. Rashin, K. Sharp, and P. Youkharibache for stimulating discussions and suggestions. This work was supported by grants from the NIH (GM-30518) and ONR (N00014-86-K-0483).

REFERENCES

1. Gilson, M.K., Rashin, A., Fine, R., Honig, B. On the calculation of electrostatic interactions in proteins. J. Mol. Biol. 183:503–516,1985.
2. Paul, C.H., Building models of globular protein molecules from their amino acid sequences. I. Theory. J. Mol. Biol. 155:53–62, 1982.
3. Rashin, A.A., Honig, B. H. On the environment of ionizable groups in globular proteins. J. Mol. Biol. 173:515–521, 1984.
4. Warwicker, J., Watson, H.C. Calculation of electric potential in the active site cleft due to α-helix dipoles. J. Mol. Biol. 157:671–679, 1982.
5. Rogers, N.K., Moore, G.R., Sternberg, M.J.E. Electrostatic

interactions in globular proteins: Calculations of the pH dependence of the redox potential of cytochrome c551. J. Mol. Biol. 182:613–616, 1985.
6. Klapper, I. Hagstrom, R., Fine, R., Sharp, K., Honig, B. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. Proteins 1:47–59,1986.
7. Warwicker, J. Continuum dielectric modelling of the protein-solvent system, and calculation of the long-range electrostatic field of the enzyme phosphoglycerate mutase. J. Theor. Biol. 121:199–210, 1986.
8. Gilson, M.K., Honig, B.H. Calculation of electrostatic potentials in an enzyme active site. Nature 330:84–86, 1987.
9. Gilson, M.K., Honig, B.H. The energetics of charge-charge interactions in proteins. Proteins 3:32–52, 1988.
10. Bockris, J.O'M., Reddy, A.K.N. "Modern Electrochemistry," v.1. New York: Plenum Press, 1977.
11. Born, M. Volumes and heats of hydration of ions. Z. Physik. 1:45–48, 1920.
12. Rashin, A.A., Honig, B. Reevaluation of the Born model of ion hydration. J. Chem. Phys. 89:5588–5593, 1985.
13. Rashin, A.A., Namboodiri, K. A simple method for the calculation of hydration enthalpies of polar molecules with arbitrary shapes. J. Phys. Chem. 91:6003–6012, 1987.
14. Gilson, M.K., Sharp, K.A., Honig, B.H. Calculating electrostatic interactions in bio-molecules: Method and error assessment. J. Comput. Chem. 9:327–335, 1988.
15. McAllister, D., Smith, J.R., Diserens, N.J. "Computer Modeling in Electrostatics." New York: Research Studies Press, Wiley, 1985.
16. McCammon, J.A., Wolynes, P.G., Karplus, M. Picosecond dynamics of tryosine side chains in proteins. Biochemistry 18:927–942, 1979.
17. Weast, R.C. (ed.). *"CRC Handbook of Chemistry and Physics"*, 56th ed. Cleveland: CRC Press, 1975.
18. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A Program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4:187–217, 1983.
19. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M.J. The protein data bank: A computer-based archival file for molecular structures. J. Mol. Biol. 112:535–542, 1977.
20. Ploegman, J.H., Drent, G., Kalk, K.H., Hol, W.G.J. Structure of bovine liver rhodanese. I. Structure determination at 2.5 Å resolution and a comparison of the conformation and sequence of its two domains. J. Mol. Biol. 123:557–594, 1978.
21. Hendrickson, W.A., Teeter, M.M. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. Nature 290:107, 1981.
22. Stenkamp, R.E., Sieker, L.C., Jensen, L.H. Binuclear iron complexes in methemerythrin and azidomethemerythrin at 2.0 Å resolution. J. Am. Chem. Soc. 106:618, 1984.
23. Satow, Y., Cohen, G.H., Padlan, E.A., Davies, D.R. Phosphocholine binding immunoglobulin Fab MCPC603: An X-ray diffraction study at 2.7 Å J. Mol. Biol. 190:593–604, 1986.
24. Novotny, J., Rashin, A., Bruccoleri, R.E. Criteria that distinguish between native proteins and incorrectly folded models. Proteins 4:19–30. (See paper in this issue.)
25. Bruccoleri, R.E., Karplus, M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. Biopolymers 26:137–168, 1987.
26. Novotny, J., Bruccoleri, R., Karplus, M. An analysis of incorrectly folded protein models: implications for structure predictions. J. Mol. Biol. 177:787–818, 1984.
27. Tanford, C., Kirkwood, J.G. Theory of protein titration curves. I. General equations for impenetrable spheres. J. Am. Chem. Soc. 79:5333–5339, 1957.
28. Desnoyers, J.E., Jolicoeur, C. Hydration effects and thermodynamic properties of ions. Mod. Aspects Electrochem. 5:1, 1969.
29. Marcus, Y. "Ion solvation." New York: John Wiley & Sons, Ltd., 1985.
30. Bash, P., Singh, U.C., Langridge, R., Kollman, P. Free energy calculations by computer simulation. Science 236:564–568, 1987.
31. Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. Science 221:709, 1983.
32. Pflugrath, J.W., Quiocho, F.A. Sulphate sequestered in the sulphate-binding protein of Salmonella typhimurium is bound solely by hydrogen bonds. Nature 314:257–260, 1985.
33. Kang, Y.K., Nemethy, G., Scheraga, H.A. Free energies of hydration of solute molecules. 3. Application of the hydration shell model to charged organic molecules. J. Phys. Chem. 91:4118–4120, 1987.
34. Eisenberg, D., McLachlan, A.D. Solvation energy in protein folding and binding. Nature 319:199–203, 1986.
35. Moult, J., James, M. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. Proteins 1:146–163, 1986.
36. Friedman, H.L., Krishnan, C.V. Thermodynamics of ion hydration. In: *"Water: A Comprehensive Treatise."* Franks, F., ed. New York: Plenum Press, vol. 3, Chapter 1, 1973.
37. Randles, J.E.B. The real hydration energies of ions. Trans. Farad. Soc. 52:1573–1581, 1956.
38. Rosseinsky, D.R. Electrode potentials and hydration energies. Theories and correlations. Chem. Rev. 65:467–490, 1965.
39. Taft, R.W., Protonic acidities and basicities in the gas phase and in solution: Substituent and solvent effects. Prog. Org. Chem. 14:247–350, 1983.
40. Wolfenden, R., Andersson, L., Cullis, P.M., Southgate, C.C.B. Affinities of amino acid side chains for solvent water. Biochemistry 20:849–855, 1981.
41. Hine, J., Mookerjee, P.K. The intrinsic hydrophilic character of organic compounds. Correlations in terms of structural contributions. J. Org. Chem. 40:292–298, 1975.