The Waters of Life

by CHAO ZHANG, S. ROY KIMURA, ZHIPING WENG, SANDOR VAJDA, RICHARD C. BROWER and CHARLES DELISI*

Biomedical Engineering Department, Boston University College of Engineering, 44 Cummington Street, Boston, MA 02215, U.S.A.

(Received in final form 13 November 1996; accepted 18 February 1997)

ABSTRACT: Water serves as a critical solvent for a remarkable array of molecules; in particular it profoundly influences the structure and activity of proteins, and their molecular interactions. Our ability to understand biological processes and to develop innovative applications for biotechnology depend in large part on understanding the biophysics of proteins in their solvated environment. The main difficulty in understanding solvation phenomena arises from the effects of electrostatics in complex biomolecular systems. In this paper we survey and critique the fundamental concepts and methodologies in evaluating electrostatic contributions to solvation. © 1997 The Franklin Institute. Published by Elsevier Science Ltd

1. Introduction

The elucidation of the electronic structure of atoms and the development of quantum theory established the electrostatic origin of all intermolecular forces (1). This is encapsulated in the Hellmen–Feynman theorem, which states that once electron distributions have been determined from the solution of the Schrödinger equation, intermolecular forces can be calculated using classical electrostatics. However, massive computations are required for ab initio molecular orbital calculations and with today’s computer technology they generally cannot be carried out for molecules with more than a few tens of atoms. The number of atoms in biological macromolecules generally exceeds this limit so that even calculations performed in vacuo would be prohibitive. The demands are even more apparent when one realizes that predicting structure requires sampling a large number of conformations in a search for the structure of minimum free energy, and the electron distribution would have to be calculated for each. The need to take account of solvation by water further compounds the problem (2).

The difficulty in calculating the internal energy of a specified geometry is circumvented by two fundamental approximations, both of which have been tested in a large number of systems, and both of which have important but limited domains of validity. First, the intramolecular calculation (solvent absent) is carried out by replacing electronic distributions by atomic charges, usually located at the positions of the nuclei.

*To whom all correspondence should be addressed [e-mail: delisi@enga.bu.edu].
Higher-order terms (i.e., dipoles and possibly quadrupoles, as discussed below) can also be used, but in most cases the approximation is truncated at the charge terms. Secondly, the interatomic forces are divided into categories, each represented by a separate algebraic form in spite of their common quantum mechanical origin. In particular, the covalent interactions are treated as mechanical constraints in terms of bond lengths, bond angles, and force constants, whereas the short range repulsion is modeled by an effective van der Waals radius. This so-called molecular mechanics approach can be used to evaluate the internal energy of any conformation of a macromolecule. Of course, all the parameters in the proposed effective potential energy function for molecular mechanics must ultimately be determined by prior ab initio calculations, supplemented perhaps by experimental data (3–8).

Determination of the structure of a large molecule (proteins in this paper) requires evaluation of free energy, not just energy. More precisely, it requires the free energy differences between different conformations of the system, protein plus water (9–11). The free energy of any conformation must be written relative to some reference state, which is often taken as the fully denatured form of the molecule (12). Then the free energy change consists of the change in conformational energy and entropy and the change in solvation energy and entropy (11–14). The conformational energy change is evaluated using the molecular mechanics approach (15) and the conformational entropy change is evaluated using an appropriate model for the backbone and side chains of the protein (16, 17). The entropy change of solvation is evaluated semi-empirically, and is generally assumed to be proportional to the change in the hydrated surface area of the molecules (12, 18, 19). The change in solvation energy will be central to this paper, the dominant effect being electrostatic, and involving the desolvation of charged groups. There is an important assumption implicit in the above discussion, namely that the individual contributions are additive (15). In molecular mechanics, additivity is generally assumed for the internal energy, but it is only an approximation for solvation contributions.

As we shall see, the effect of water makes the determination of the electrostatic contribution to the free energy of such processes as protein folding and protein ligand association very difficult. It is, however, central to these processes and to understanding biomolecular structures, their interactions and design. Indeed, as implied by W.H. Auden's poetic phrase which we used as the title, the effort to understand solvation is the biophysicist's equivalent of the search for the holy grail or for the fountain of youth. It is one of the grand quests of modern science.

2. The Microscopic Picture of Protein Electrostatics

2.1. The electrostatic fields and energies

The electrons and nucleus of a monoatomic molecule define a charge distribution, \( g(r') \), where \( r' \) locates charge on the molecule. The electric field of the charge distribution is given by

\[
e(r) = \int \frac{g(r')(r-r')}{|r-r'|^3} \, dv',
\]

(1)

where \( r \) is the point of observation outside the molecule.
For distances \( r = |r| \) that are large compared to the dimensions of the molecule, we can expand Eq (1), and thereby approximate the field by the electric field of a point charge \( q \) and a dipole \( \mathbf{p} \) located at the center of the molecule

\[
e(r) \simeq q r / r^3 - (p - 3(r \cdot p)r/r^3)/r^3,
\]

where \( q = \int q(r') \, dv' \) and \( p = \int q(r') r' \, dv' \). With this representation, the electrostatic potential is given by

\[
\phi(r) \simeq q/r + p \cdot r/r^3.
\]

Dipole moments originate from a shift in the centers of mass of positive and negative charges due to either permanent electron displacement (permanent dipole) or instantaneous electron fluctuation (induced dipole). Both effects are very important for macromolecular stability. In principle the quadrupole and higher-order terms are also present, but their effects are normally negligible.

For a system consisting of a collection of such molecules, the electrons and nuclei can be expanded into point charges \( q_i \) and dipoles \( \mathbf{p}_i \), located at the centers of individual molecules \( r_i \), from a common origin, where \( i = 1, 2, ..., n \). The total potential energy of the system is

\[
W = \frac{1}{2} \sum_{i=1}^{n} (q_i \phi(r_i) - \mathbf{p}_i \cdot \mathbf{e}(r_i)),
\]

where \( \phi(r_i) \) and \( \mathbf{e}(r_i) \) are, respectively, the electrostatic potential and field at \( r_i \), exerted by all other charges and dipoles. The factor 1/2 is introduced since each interaction is counted twice while summing over all the charges and dipoles. Most chemical energies are expressed in kcal/mol. \( W \) can be converted into kcal/mol units by multiplying by the constant, 332.0. The units used throughout this paper are distance, Å; charge, in multiples of electrons; and energy in kcal/mol.

For a molecule that consists of covalently bonded atoms, each atom possesses a charge distribution that can be expanded to a point charge and a dipole at the center of the atom. As a result, the charge distribution of the molecule is represented by a collection of point charges and dipoles. The intramolecular electrostatic interactions between non-bonded atoms as well as intermolecular interactions in a multi-molecular system can still be calculated using Eq (4). However, for two atoms that are linked by a covalent bond, or connected through an intermediate atom, or restricted via a dihedral angle, the interatomic interaction is often treated as a mechanical constraint rather than a charge-charge interaction. In this case, bond energy, angular energy, and dihedral energy are defined using a spring model and various force parameters.

This paper focuses on systems containing solvated proteins; both components, protein and water molecules, having covalent structures. The electrostatic interactions between protein and water molecules and the intramolecular interactions among non-bonded protein atoms are our main interest. We now elaborate on the sources of the various terms in Eq (4) in the highly heterogeneous protein environment.

2.2. The electronic properties of proteins

2.2.1. Ionizable groups. At neutral pH, isolated amino acids are zwitterions (NH\(^+\)-C\(_\alpha\)(R)-COO\(^-\)) \( \text{(20)} \), the amino group being positively charged and the carboxyl group negatively charged. When amino acids are linked in a protein, the backbone is elec-
trically neutral, with the amino terminus having a positive charge, and the carboxyl terminus having a negative charge. Certain amino acid side chains can also have net charges. In particular, the side chains of aspartic acid (Asp) and glutamic acid (Glu) are weak acids, and tend to 'donate' protons to solution and acquire unit negative charge, whereas the side chains of lysine (Lys) and arginine (Arg) are weak bases, and tend to attract protons from solution and acquire unit positive charge.

The actual charge state of acidic and basic side chains is governed by the affinity of protons for the group. If the hydrogen ion (H\(^+\)) concentration, or pH\(_a\), is varied, the point at which 50% of the ionizable groups of a given type are charged is referred to as its pK\(_a\). The side chain pK\(_a\)s of Asp, Glu, Lys and Arg in dilute solution are 3.9, 4.3, 10.5 and 12.5, respectively (20). Therefore, the first two are negatively charged at neutral pH, while the last two are positively charged.

Molecular mechanics simulations usually assume that each ionizable group exists in a single fixed charge state whose selection is based upon the pK\(_a\) of the isolated group in solution, rather than in its actual protein site. However, the association of H\(^+\) with an ionizable group will be influenced by the local electrostatic field (21-23), and that in turn depends on the spatial distribution of other charges in the protein, i.e. on the protein's conformation. Such field-induced shifts on pK\(_a\)s, sometimes by as much as 3 pH units (24), can affect the equilibrium conformation of a protein. The consequences can be particularly dramatic for histidine (His), whose side chain pK\(_a\) of 6-6.5 is in the biologically relevant pH range. In some proteins, His can act as a switch, with its charge state switching with changes in environment (biological compartments vary slightly in their pH) and inducing transitions between protein conformations, or influencing the ability of proteins to pair (25, 26).

2.2.2. Partial charges and permanent dipoles. Two atoms that are covalently bonded have overlapping electron orbitals. Although the atoms are said to share electrons, the distribution of electrons in the vicinity of the two nuclei is asymmetric, the electrons having a higher probability of being near atoms of greater electronegativity. The variation in electronegativity for various non-metallic atoms of biological interest—O(3.45), N(2.98), C(2.55), S(2.53), H(2.13) and P(2.10) (27)—reflects the structure of their outer orbitals. The asymmetric electron distribution attending covalent bond formation can be approximated by a pair of partial electronic charges, separated by the bond length, the charge \(\delta^+\) (\(\delta^-\)) being associated with the least (most) electronegative atom (27). Such a polar group has a permanent electric dipole, and its orientation is influenced by the local field as well as steric restraints.

Traditionally, the magnitudes of partial electronic charges are obtained by fitting analytic potential functions to quantum mechanical calculations of well-defined model systems and to relevant experimental information (3, 4, 28, 29). The two most widely-used force fields for protein energy and molecular dynamics calculations are the CHARMM force field developed by Karplus and co-workers (5) and the AMBER force field developed by Kollman and co-workers (6). Both provide a self-consistent set of partial electronic charges for each of the twenty naturally occurring amino acids. Recently, partial charges have also been obtained by fitting an empirical electrostatic model to the solvation free energies of amino acids (7, 8). Although the differences between various sets of partial charges appear small, they can lead to markedly different results in some applications.
A weak bond between two electronegative atoms such as nitrogen and oxygen can form if they share a hydrogen atom, the hydrogen atom typically being covalently attached to the donor (27). This sharing leaves the hydrogen with a partial positive charge, and both the donor and acceptor with partial negative charges. Such hydrogen bonds are strongly directional (30, 31), and range in strength from about 3 to 7 kcal/mol in vacuum (27), this being intermediate between weak van der Waals interactions (0.3 kcal/mol) and covalent chemical bonds (100 kcal/mol). Energetically, the hydrogen bond is sometimes described as the interaction between two permanent dipoles, capturing the well-known high directionality of the bond, although it is said that hydrogen bonds also have some 'covalent' character.

The aromatic side chain atoms of phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp) are called weakly polar atoms (32). The benzene-like aromatic rings of these amino acids demonstrate a characteristic separation of partial charges. The presence of double bonds between ring carbon atoms gives rise to a $\delta^-$ $\pi$-electron* cloud covering the face of the aromatic ring and $\delta^+$ hydrogen atoms bound to each of the ring carbon atoms (32). The typical partial electronic charge values are about $\pm 0.15$ electrons in the vicinity of each hydrogen nucleus and each carbon nucleus, respectively (5, 6).

The long-range electrostatic effect of aromatic side chains is very small. Take benzene as an example. Since the arrangement of partial charges in benzene consists of three simple quadrupoles, the distance dependence of the oxygen–aromatic interaction varies as $1/r^6$ for an uncharged oxygen atom and $1/r^3$ for carboxylate oxygen. However, at short distance, the charge separation in aromatic groups permits polar or charged groups to approach only along certain directions, as observed in both protein (33) and small-molecule crystal structures (34). Specifically, negatively charged oxygen atoms ($\delta^-$) are preferentially found in the plane of the aromatic ring near the $\delta^+$ hydrogen atoms (35, 36), whereas the side-chain amino groups of Lys, Arg, glutamine (Gln), asparagine (Asn) and His tend to appear axially near the $\delta^-$ $\pi$-electron clouds of aromatic rings and avoid equatorial positions near the $\delta^+$ hydrogen atoms of these rings (37). These distributions are different from what are expected from random close packing of groups within the hydrophobic core of a protein (32), indicating that such oxygen– and amino–aromatic interactions in proteins are involved in stabilizing protein structures.

As opposed to the face-to-face $\pi$–$\pi$ stacking of the rings in DNA, the aromatic–aromatic interactions in proteins prefer to associate via enthalpically favorable, edge-to-face interactions in which a $\delta^+$ hydrogen atom from one ring makes a close contact with the $\delta^-$ $\pi$-electron cloud of the other ring (32). Overall, the aromatic–aromatic interaction has a distance dependence that varies as $1/r^6$ and is spatially anisotropic.

2.2.3. Electronic polarizability and induced dipoles. In most cases the charge distribution in a collection of molecules depends on the field that each molecule applies to its neighbors. Polarizability is defined conceptually as the propensity for an electron

*Each carbon in the benzene ring forms three coplanar $sp^2$-hybrid $\sigma$ bonds at 120° angles. These carbon–carbon, carbon–hydrogen $\sigma$ bonds use three of the four valence electrons of each carbon. The remaining six carbon electrons are in parallel $p$ orbitals, one on each of the six carbons. These electrons are collectively called $\pi$ electrons.
distribution to change shape and acquire an induced dipole in response to an applied field. Specifically, when a polarizable group $k$ is located at $r_k$, a dipole

$$p_k = \alpha_k e(r_k)$$

(5)

is induced by the local field $e(r_k)$, where $\alpha_k$ is the polarizability constant of the group. Evidently, highly polarizable atoms will tend to have low electronegativities. Thus the polarizability of CH$_3$ (1.97) is greater than that of NH$_3$ (1.44) or OH (0.73). Polarizability increases with atomic radius, since electron binding energy decreases with increasing distance from the nucleus.

The total energy for a collection of charges ($q_i$ at $r_i$) and induced dipoles ($p_k$ at $r_k$) is given by

$$W = \frac{1}{2} \left\{ \sum_i q_i \phi(r_i) - \sum_k p_k \cdot e(r_k) + \sum_k \alpha_k |e(r_k)|^2 \right\},$$

(6)

where the first two terms are the same as those in Eq (4) and the last term represents the energy invested in forming the induced dipoles (22).

Consider a simple system consisting of one charge ($q_i$) and one polarizable group (polarizability constant $\alpha_k$). Denote $r = r_k - r_i$, then the field at the polarizable group produced by $q_i$ is $e(r_k) = q_i (r/r^3)$. Consequently, the dipole induced at the polarizable group will be $p_k = \alpha_k q_i (r/r^3)$. The first two terms in Eq (6) give an energy of $-\alpha_k q_i^2/r^4$. The energy required for forming the dipole is $\frac{1}{2} \alpha_k e(r_k) \cdot e(r_k) = \frac{1}{2} \alpha_k q_i^2/r^4$. Therefore, the total potential energy is $-\frac{1}{2} \alpha_k q_i^2/r^4$, i.e. half of the value estimated if the energy for forming the dipole is not considered.

The above relation can be generalized. Using the relation $p_k = \alpha_k e(r_k)$, Eq (6) can be further simplified as (22)

$$W = \frac{1}{2} \left\{ \sum_i \sum_j q_i q_j / r_{ij} - \sum_k p_k \cdot e^0(r_k) \right\},$$

(7)

where $e^0(r_k)$ is the field on the $k$th dipole from the point charges only.

It is evident from the above that at short distances, electron distortion effects are important. The polarization effect is generally difficult to introduce in force-field simulations owing to the use of a fixed set of parameters to describe the molecule in different environments. However, over the past 20 years, serious attempts to model proteins using experimentally determined polarizabilities have been successfully carried out by Warshel and coworkers (see Section 2.4).

Finally, even if two atoms are neutral and have no permanent dipoles, they will be attracted to one another by a short-range force. The basis for such van der Waals attraction is that the distribution of electronic charge around an atom fluctuates rapidly. The transient asymmetry in the electronic charge around an atom induces a transient asymmetry in the electron distribution around its neighboring atoms. As the distance between the two atoms, $r$, decreases, the attraction energy between the induced dipoles increases as $r^{-6}$. The increase continues until the atoms come so close that the inner electron clouds of the two atoms interact with each other and a strong repulsive force comes into play. The magnitude of repulsive energy is generally modeled adequately with an $r^{-12}$ potential.
2.3. The orientational polarization of polar solvent

The second component in protein solutions is water (H₂O). The oxygen in a water molecule has a partial negative charge owing to its electronegativity and the two attached hydrogen atoms have partial positive charges. These atoms are arranged in a non-linear manner that makes the water molecule ideally suited to engage in a hydrogen bonding network. Liquid water thus has an ordered structure in which hydrogen-bonded clusters of molecules are continually forming and breaking up (38, 39).

When a charged molecule is dissolved in such a polar liquid, nearby solvent molecules will rotate in response to the field, so as to decrease the free energy of the system. These oriented solvent molecules produce electric fields of their own, which oppose the fields produced by the charged group of the solute. In addition, this orientation effect is often subject to the steric restraints of the system (e.g. van der Waals interactions and other intermolecular forces).

Warshel and coworkers suggest that the main physical effect of solvation can be captured by the interactions between the dipoles of many solvent molecules and the solute charges (22). If the energetics of solvation rather than the precise structure of water is of major interest, water molecules can be represented by permanent dipoles. In such a model, solvent molecules interact with each other by dipole–dipole and van der Waals interactions and by additional correction potentials for short-range interactions. Thus, for a system consisting of charges (\(q_i\)) and polar solvent molecules represented by permanent dipoles (\(\mathbf{p}_k\)), the total potential energy can be expressed as

\[
W = \frac{1}{2} \left\{ \sum_i q_i \phi(r_i) - \sum_k \mathbf{p}_k \cdot \mathbf{e}(r_i) \right\} + \sum_k U(r_i, \mathbf{p}_k).
\]

where \(U(r_i, \mathbf{p}_k)\) is the effective energy for placing the \(k\)th dipole at its equilibrium position and orientation (assuming that the solute molecules are not affected). Thus \(U(r_i, \mathbf{p}_k)\) represents the energy contributions associated with building up the final configurations of solvent dipoles that are not explicitly included in the first two terms. Note that in this case dipoles have the freedom to move, so \(U(r_i, \mathbf{p}_k)\) has other origins apart from electrostatics.

When the electric field exerted by solute charges on the solvent (\(\mathbf{e}\)) is small, the orientation effect for the \(k\)th dipole due to \(\mathbf{e}(r_k)\) can be represented by a small perturbation, \(\Delta \mathbf{p}_k\), relative to its equilibrium configuration, \(\mathbf{p}^0\). In this case, an approximate expression for \(W\) is given by (22):

\[
W = \frac{1}{2} \left\{ \sum_i \sum_{j \neq i} q_i q_j / r_{ij} - \sum_k \Delta \mathbf{p}_k \cdot \mathbf{e}(r_k) \right\}.
\]

In addition to solvent water molecules, electrostatic interactions in proteins are shielded by the presence of small ions, typically dissolved salts. Positive ions tend to be located near the negative charges of proteins, and negative ions near the positive protein charges. For very dilute solutions, the electrostatic effects outside the protein can be described by the Debye–Hückel theory (see Section 3.1).

2.4. Microscopic models: embracing the details

Detailed modeling of electrostatic effects in proteins requires consideration of all charges and permanent dipoles of both the protein and the solvent, plus their atomic
polarizabilities. This leads to the so-called microscopic models which consider the detailed pairwise interactions between all permanent and induced charges in the entire system and treat them individually.

In principle, brute-force simulations can be carried out for a large box that contains the macromolecule and explicit water molecules which interact with one another as well as with the macromolecule. The main difficulty with this approach is that the computational cost is prohibitive in most applications. In fact, in addition to having to include large numbers of water molecules (usually several thousand) in order to accurately account for the long-range electrostatic effects, it is also necessary to sample over a range of thermodynamic variables with long autocorrelation times to compute free energy differences by thermodynamic integration or free energy perturbation methods (9, 10). Such calculations are prohibitive in applications that require repeated evaluation of the free energy as in a conformational search, or for a number of different molecules as in drug design.

One way of obtaining reasonable results with a limited number of solvent molecules is to use spherical boundary conditions with special surface constraints (40, 41). This method employs an all-atom representation for both the macromolecule and the water molecules and places them in a sphere. The basic idea is that proper constraints imposed on the spherical boundary can force the finite system to behave as the corresponding region in an infinite system (40, 22, 41). In addition, the long-range electrostatic effects associated with the region outside the sphere are reproduced by a dielectric continuum. Such a hybrid strategy allows for the accurate microscopic representation of water in applications where this is important, while significantly saving computation by reducing the number of particles in the simulation. However, as pointed out by (41), this method suffers from two problems: the time delay in the response of the reaction field and the instability of the total dipole moment, due to the positive feedback between this quantity and the reaction field.

The ordinary potential functions of molecular mechanics treat partial atomic charges as fixed, and approximate the electronic polarizability of the solute and solvent with a uniform dielectric constant. To overcome this limitation, Warshel and Levitt (28) introduced a model that explicitly considers the differences in electronic polarizability of individual atoms within proteins. They treat protein atoms as bearing a point charge and an induced dipole. The magnitude and direction of each dipole is determined from the standard electrostatic relationship (Eq (5)). The field acting at a given atom is calculated as the sum of the field due to fixed charges and that arising from all induced dipoles on other atoms in an iterative scheme. To overcome the computational burden imposed by full treatment of all-atom models of liquid, Warshel and Levitt (28) also use a Langevin dipole model for water molecules that are beyond an explicit boundary solvation layer. The so-called protein dipole–Langevin dipole (PDLD) model makes it possible to treat electrostatic contributions to enzyme catalysis in a physically rigorous fashion for the first time (21, 22, 42–44).

An alternative approach to electrostatic interactions is based on the classical theory of continuum electrostatics. This approach specifies the macromolecule in terms of a set of atomic charges at discrete positions in a low dielectric medium, which is surrounded by a high dielectric medium which represents the mean effects of water. Proper matching of boundary conditions between the molecular and continuum description
may in principle provide an approximate expression for the field everywhere. The advantage of the continuum theory is that it can readily describe the bulk properties of molecules, although very accurate electrostatic energies cannot be provided.

3. Continuum electrostatic models

3.1. Electrostatics in the continuum

3.1.1. Basic concepts and Poisson's equation. Continuum electrostatic treatments are based on the definition of macroscopic averages. The macroscopic electric field, \( \mathbf{E} \), is defined as the average of microscopic electric field, \( \mathbf{e} \), in a volume element of the system. Similarly, for a collection of microscopic dipoles \( \mathbf{p} \), the corresponding macroscopic polarization \( \mathbf{P} \) is defined as the volume average of \( \mathbf{p} \). The quantities \( \mathbf{E} \) and \( \mathbf{P} \) are related to each other by \( \mathbf{E} = \mathbf{D} - 4\pi \mathbf{P} \), where \( \mathbf{D} \), 'the electric displacement vector', is defined as the macroscopic average of the electric field due to free charges in the absence of polarizable medium (i.e. dielectric media).

Macromolecular systems are normally treated as linear isotropic media where \( \mathbf{D} \) is assumed to be parallel to \( \mathbf{E} \). Under this condition, the electric displacement and the macroscopic electric field are related by \( \mathbf{D} = \varepsilon \mathbf{E} \), where \( \varepsilon \), the dielectric constant, represents the screening of the external field by the polarization of a volume element (45). For example, water diminishes the strength of the electrostatic field by a factor of 78.5 (at 25°C), the dielectric constant of water, compared with the same field in a vacuum.

In a homogeneous dielectric medium, the potential due to free charges can be written as

\[
\Phi(r) = \frac{1}{\varepsilon} \int \rho^f(r') \frac{\text{d}v'}{|r-r'|} \tag{10}
\]

where \( \rho^f \) is the volume density of the free charges. Here we use \( \Phi \) to indicate the macroscopic definition of potential, distinguishing it from the microscopic definition, \( \phi \). By definition, \( \Phi \) contains contributions from all charges, including the potential from a charge to itself.

In general, the so-called 'dielectric constant' is not a constant but varies from location to location in a heterogeneous system. The simple expression (10) does not hold for such general cases and Poisson's equation (of which Eq (10) is the solution for a special case) must be solved explicitly:

\[-\nabla \cdot \varepsilon(r) \nabla \Phi(r) = 4\pi \rho^f(r) \tag{11}\]

where \( \Phi(r) \), \( \varepsilon(r) \), and \( \rho^f(r) \) are, respectively, the electric potential, dielectric constant, and free charge density at \( r \).

3.1.2. Electrostatic energies and reaction fields. A general understanding of electrostatic energetics comes from the classical theory of the interactions between charges and continuum dielectrics. This theory represents dielectric effects by the bound or polarization charges induced by the electric field of free charges. Accordingly, the total charge distribution can be written as

\[
\rho^t(r) = \rho^f(r) + \rho^b(r), \tag{12}\]

where \( \rho^b \) is the volume density of the bound charges.
By definition, the total energy of an electrostatic system \( W \) is the work done on the system in carrying its elements of free charge from infinity to the specified distribution

\[
W = \frac{1}{2} \int \rho_s(r) \Phi(r) \, dv
\]  
(13)

where \( \Phi \) is the electrostatic potential due to all charges both free and bound. Separate the terms,

\[
W = \frac{1}{2} \int \frac{\rho_s(r) \rho_s(r') \, dv \, dv'}{|r-r'|} + \frac{1}{2} \int \frac{\rho_s(r) \rho_b(r') \, dv \, dv'}{|r-r'|} .
\]  
(14)

The first term corresponds to the energy of the system of free charges in the absence of dielectrics, and the second term corresponds to the energy of the interaction of free charges with the bound charges induced in the dielectrics by these charges. The potential due to the bound charges is often called the potential of the 'reaction field'.

3.1.3. \textit{Dilute ion clouds}. Ionized salts at physiological concentration also play a role in shielding interactions between charged groups on the surface of a protein. Quantitative theories for biologically relevant ionic strengths (\( \approx 0.15\text{mM} \)) have been difficult to develop. An important analytical result which captures the shielding effect qualitatively, and is also quantitatively accurate for very dilute ionic solutions, was developed by Debye and Hückel (46). In general, the charge density at a distance \( r \) from a spherical central ion is an exponential function of the pairwise potential of mean force. The central approximation of Debye and Hückel replaces the potential of mean force by \( q\Phi(r) \). This is valid only for dilute ionic solutions since the 'ionic atmosphere' around the ion of interest, which normally follows a Boltzmann distribution, will generally be distorted in the presence of other ions nearby which carry their own ionic atmospheres. The approximation leads to the so-called Poisson–Boltzmann (PB) equation,

\[
-\nabla \cdot \varepsilon(r) \nabla \Phi(r) = 4\pi \rho(r) - \varepsilon(r) \kappa^2 \sinh(\Phi(r)).
\]  
(15)

Here, \( \kappa = \sqrt{8\pi \alpha N e^2 I/(\epsilon_w k_B T)} \left( = \sqrt{I/3.04 \text{Å}^{-1}} \right. \) at 25°C, \( e \) is the charge of an electron, \( N \) is Avogadro’s number, and \( I \) is the ionic strength in units of millimoles per liter. The quantity \( 1/\kappa \), known as the Debye length, is the distance at which screening by mobile ions reduces the (linearized) potential by \( e^{-1} \).

In their original paper, Debye and Hückel proceed by Taylor expanding the non-linear Boltzmann term and neglecting everything beyond the linear term in the expansion, making the assumption that \( e\Phi/k_B T < 1 \) (46). The resulting linearized Poisson–Boltzmann (LPB) equation is formally known as the Helmholtz equation

\[
-\nabla \cdot \varepsilon(r) \nabla \Phi(r) = 4\pi \rho(r) - \varepsilon(r) \kappa^2 \Phi(r).
\]  
(16)

3.2. \textit{Electrostatic solvation effects}

3.2.1. \textit{Cavity model and dielectric discontinuity}. A fundamental continuum model of a protein–solvent system treats water and protein as continuous media with different dielectric constants; the dielectric discontinuity is defined by the surface of the protein molecule. The homogeneous solvent region has the same dielectric constant as pure liquid water (\( \epsilon_w = 78.5 \)). The cavity which the protein occupies is assumed to have dielectric constant \( \epsilon_p \) which reflects the protein’s own ability to polarize. Since the polarizability of the interior of proteins is inhomogeneous, the proper choice of internal
dielectric constant is not obvious and necessarily involves a gross approximation. We will discuss this issue in Section 3.5.

With this representation, the protein molecule is treated as a collection of atom-centered point charges inside a cavity with dielectric constant $\varepsilon_p$, surrounded by an external continuum with dielectric constant $\varepsilon_w$. For such a system, Eqs (11) and (15) can be solved in each region using a uniform dielectric constant. The key to a rigorous formulation is having the appropriate boundary condition between the two regions.

According to the theory of dielectrics, the electric field of the solute charges in the presence of a dielectric interface is equivalent to the linear superposition of the field of the solute charges in a uniform medium of dielectric $\varepsilon_p$ and the field of an appropriate surface charge distribution ($\sigma$) on a surface of the same shape as the dielectric interface, but in a vacuum environment. The total electrostatic energy, $W$, of assembling the solute charges in the cavity from infinite separation is

$$W = \frac{1}{2} \sum_i q_i \Phi_i = \frac{1}{2} \sum_i q_i \left(\Phi_i^f + \Phi_i^b\right),$$

(17)

where the summation is over the solute charges and $\Phi_i$ is the potential at the position of solute charge $q_i$. $\Phi_i^f$ and $\Phi_i^b$ are, respectively, the electrostatic potentials due to the solute and bound charge distributions. $\Phi_i^f$ and $\Phi_i^b$ can be computed as

$$\Phi_i^f = \frac{1}{\varepsilon_p} \sum_k \frac{q_k}{|r_i - r_k|},$$

(18)

$$\Phi_i^b = \int \frac{\sigma(\vec{r})}{|r_i - \vec{r}|} \ dS(\vec{r}),$$

(19)

where $\sigma(\vec{r})$ is the surface charge density and the integral is taken over the entire closed boundary surface.

3.2.2. Electrostatic free energy of solvation. The term solvation refers to the surrounding of each dissolved molecule by a shell of more or less tightly bound solvent molecules. This solvent shell is the result of intermolecular forces between solute and solvent. Here, the electrostatic free energy of solvation is defined as the difference between the total electrostatic energies, Eq (17), in water and in a reference medium $\varepsilon_p$. If water is assumed to behave as a linear dielectric, then the electrostatic part of the solvation energy equals the energy of interaction of solute charges with the reaction field:

$$W^{\text{solv}} = \frac{1}{2} \sum_i q_i \Phi_i^b,$$

(20)

where $\Phi_i^b$ is the electrostatic potential of bound charges due to the dielectric discontinuity. Based on the same principle, when the solute molecule is transferred from one solvent (I) to another (II), the change in electrostatic solvation energy can be expressed as

$$\Delta W^{\text{solv}} = \frac{1}{2} \sum_i q_i \left(\Phi_i^b\right)_I - \frac{1}{2} \sum_i q_i \left(\Phi_i^b\right)_II.$$  

(21)

3.2.3. Self energies and screening effects. $\Phi_i^f$ in Eq (20) can be separated into two
terms: one representing the reaction field due to charge $i$ itself (self term), the other representing the reaction fields of all other charges (cross term):

$$\Phi^b_i = \Phi_i^{b\text{self}} + \sum_{j \neq i} \Phi_i^{b\text{cross}}.$$  

(22)

In turn, the total electrostatic energy of solvation can be rewritten as

$$W^{\text{solv}} = \sum_i W_i^{\text{self}} + \frac{1}{2} \sum_i \sum_{j \neq i} W_i^{\text{cross}}$$

$$= \frac{1}{2} \sum_i q_i \Phi_i^{b\text{self}} + \frac{1}{2} \sum_i \sum_{j \neq i} q_i \Phi_i^{b\text{cross}}.$$  

(23)

When infinitely separated charges are brought together in a homogeneous medium of dielectric $\epsilon_p$, the total energy of this process is simply the sum of pair-wise Coulombic interactions:

$$W_{ij}^{\text{Coul}} = \frac{q_i q_j}{\epsilon_p r_{ij}}.$$  

(24)

If this process occurs in the vicinity of a dielectric interface, additional interactions between these charges and the reaction field of other charges due to dielectric discontinuity come into play. This term is reflected by $W_{ij}^{\text{cross}}$. We can define the total pair-wise interaction energy between charges $i$ and $j$ as

$$W_{ij}^{\text{pair}} = W_{ij}^{\text{Coul}} + W_{ij}^{\text{cross}}.$$  

(25)

In general, $W_{ij}^{\text{cross}}$ has the opposite sign from $W_{ij}^{\text{Coul}}$. Consequently, the pair-wise electrostatic interaction between charged groups of a solute in water is markedly weakened. This screening effect can be described via the definition of the effective dielectric constant $\epsilon_{\text{eff}}$:

$$\epsilon_{\text{eff}} = \frac{q_i q_j}{r_{ij} W_{ij}^{\text{pair}}}.$$  

(26)

$\epsilon_{\text{eff}}$ simply represents all factors associated with the particular charge–charge interaction that are not included in the Coulomb formula. In most cases, $\epsilon_{\text{eff}} > \epsilon_p$. However, it should be pointed out that this screening effect does not imply a decrease in the total electrostatic energy which also includes self energies $W_i^{\text{self}}$ and $W_j^{\text{self}}$. In fact, the interactions between charges in water are not weakened but rather compensated by solvation of individual charges (22, 47). Both the self-energies of single charges and the interaction energy of the charge pair ($W_{ij}^{\text{pair}}$) depend on the geometry of the solute (48).

We use the following analytical example to illustrate these physical concepts. If the interface between two dielectric regions is a plane, the electrostatic potential ($\Phi^b$) can be determined by setting up appropriate image charges instead of solving Poisson’s equation directly. Consider a single point charge, $q_*$, embedded in a semi-infinite nonpolar dielectric medium ($\epsilon_p$) and located at a distance $d$, from a plane boundary which separates the medium from semi-infinite water (dielectric $\epsilon_w$). The electric potential in the nonpolar region can be obtained by replacing the water region with an $\epsilon_p$ medium and a point charge of $q_* = -q_*(\epsilon_w - \epsilon_p)/(\epsilon_w + \epsilon_p)$ at the same distance from the
boundary as the actual charge except on the opposite side of the boundary (i.e. the mirror image position). Then \( \Phi^{\text{self}} \) in Eq (22) is equal to the potential due to this image charge. Therefore, the (self)-solvation energy is

\[
W_{\text{self}}^s = -\frac{1}{2\epsilon_p} \frac{\xi q_i^2}{2d_i}.
\]

(27)

where \( \xi = (\epsilon_w - \epsilon_p)/(\epsilon_w + \epsilon_p) \) sets the magnitude of the image charge \( q_* = -\xi q_i \).

When a second charge, \( q_j \), is introduced into the nonpolar medium at a distance \( d_j \) from the planar boundary, the total electrostatic solvation free energy becomes

\[
W_{ij}^{\text{soln}} = -\frac{\xi}{2\epsilon_p} \left( \frac{q_i^2}{2d_i} + \frac{q_j^2}{2d_j} + \frac{q_i q_j}{d_{ij}} + \frac{q_i q_j}{d_{ij^*}} \right)
\]

(28)

due to the cross-interactions between \( q_i \) and \( q_j \)'s image, \( q_j \) (distance \( d_{ij^*} \)) and between \( q_i \) and \( q_j \)'s image, \( q_j^* \) (distance \( d_{ij} \)). The pair-wise electrostatic energy between \( i \) and \( j \) is

\[
W_{ij}^{\text{pair}} = \frac{q_i q_j}{\epsilon_p d_{ij}} - \frac{\xi q_i q_j}{\epsilon_p d_{ij^*}}.
\]

(29)

In general, models involving dielectric regions such as the above fall into the standard repertoire of problems in mathematical physics for which analytical methods have been developed extensively over the past 100 years. Despite this, analytical solutions exist only for highly symmetric, simple geometries such as half planes, spheres and ellipses. Fortunately, the recent and continuing development of sophisticated numerical methods by researchers is beginning to permit the calculation of electrostatic effects in irregular, more realistic geometries.

### 3.3. Kirkwood–Tanford model

The first macroscopic theory of protein electrostatics was developed by Kirkwood (49) and later modified by Tanford and Kirkwood (50). This model consists of an array of fixed charges inside a spherical region (radius \( a \)) with low dielectric constant \( \epsilon_p \), embedded in a high dielectric continuum (\( \epsilon_w \)) (49, 50). It also includes mobile ions which can come within a distance \( b > a \) of the center of the protein. Hence the potential field is separated into three regions: (i) the inner sphere (\( r < a \)) within which the potential is governed by Laplace’s equation (except at the point charges); (ii) the middle shell region (\( a < r < b \)) which, again, allows use of Laplace’s equation (since ions are excluded from this region); and (iii) the outer region (\( r > b \)) which contains the ions and for which we must use the (linearized) Poisson–Boltzmann equation.

The potential due to a point charge at \( r' \) in the inner region is given by

\[
\Phi = \frac{q}{\epsilon_p |r-r'|} + \Phi_{\text{cs}} + \Phi_{\text{ci}},
\]

(30)

where

\[
\Phi_{\text{cs}} = \frac{q}{a} \sum_{n=0}^{\infty} \frac{(n+1)(\epsilon_p - \epsilon_w)}{\epsilon_w(n+1) + n \epsilon_p} \left( \frac{rr'}{a^2} \right)^n P_n(\cos \theta)
\]

(31)

is the potential due to the induced polarization at the protein–solvent interface, and
\[ \Phi_{ei} = -q \left( \frac{kb}{1 + kb} + \sum_{n=1}^{\infty} \frac{2n+1}{2n-1} \left( \frac{\epsilon_w}{\epsilon_w(n+1) + n\epsilon_p} \right)^2 \right) \]

\[
\left\{ \begin{array}{l}
(\frac{r}{b})^n P_n(\cos \theta) \\
\frac{K_{n+1}(kb)}{K_{n-1}(kb)} + \frac{n(\epsilon_w - \epsilon_p)}{\epsilon_w(n+1) + n\epsilon_p} \left( \frac{a}{b} \right)^{2n+1} \frac{(kb)^2}{4n^2 - 1}
\end{array} \right\}
\]

(32)

is the potential due to the charge density of the ion distribution.

Letting \( \rho = rr' / a^2 \) and \( z = \cos \theta \), we may rewrite \( \Phi_{ei} \) in the following integral form (see Appendix):

\[
\Phi_{ei} = -\frac{\xi q}{\epsilon_p a} \frac{1}{\sqrt{1 + \rho^2 - 2\rho z}} - \frac{\xi q}{\epsilon_p + \epsilon_w a} \int_{0}^{1} x^{\delta - 1} dx,
\]

(33)

where \( \xi = (\epsilon_w - \epsilon_p) / (\epsilon_p + \epsilon_w) \) as before and \( \delta = \epsilon_w / (\epsilon_p + \epsilon_w) \).

We notice that the first term in this expression is due to an image charge in the outer dielectric located at \( r^* = a^2 r / r^2 \), with a magnitude \( q^* = -\xi q \) reduced by a factor of \( \xi \) just as in the planar case. In general, converting an infinite sum into a closed integral form allows for rapid evaluation with the use of numerical quadrature integration algorithms. In the above case, however, the dominant contribution, including the divergence at \( r = a \) has been separated out in the first term. Consequently, the integral term with an extra factor of \( \epsilon_p / (\epsilon_p + \epsilon_w) \approx 0.02 \) in Eq (33) contributes less than two percent to the total potential.

Based on Eq (23), the self-energy of the charge is \( 1/2q^2 \Phi_{ei} \). When the charge is located at the center of the sphere, the self-energy becomes the Born energy (51):

\[
W^{\text{Born}} = \frac{q^2}{2a} \left( \frac{1}{\epsilon_w} - \frac{1}{\epsilon_p} \right).
\]

(34)

This equation gives the work required to grow a point charge (or equivalently, to grow any spherically symmetric charge distribution) inside a spherical cavity with given radius \( a \) embedded in a dielectric (\( \epsilon_w \)). The reduction in energy captures the basis for the ease of solvating ions in water. Equation (34) predicts, with a fair degree of success, the hydration free energies (\( \Delta G \)) of simple ions in water given the charge (\( q \)) and radius (\( a \)) of the ion.

The exact geometry of proteins, such as the detailed convolutions of their surfaces, significantly affects the electrostatic contribution to their free energies. This, in turn, can largely alter predictions of their functional properties, pointing to the need for numerical methods. These methods are discussed in the following section.

3.4. Numeric solutions

3.4.1. Finite difference and finite element methods. With the aid of high-speed computational techniques, the exact shapes of proteins can be considered by solving Poisson's equations numerically. Historically, Ortug (52) first moved beyond simple
geometries by exploiting the finite element method for solving Poisson's equations. This method divides the continuum space into a collection of subregions, called 'finite elements', and expresses the solution in a typical element in terms of simple interpolation functions and the physical parameters at the nodes of the element. The relationships between elements and global representations of node parameters are assembled into a large system of linear equations; the coefficients obtained by solving these equations are the potentials at the nodes. The finite element method permits the description of systems with arbitrary geometry with high accuracy and flexibility (52, 53). Grids of mixed coarseness can be used to obtain an accurate description of the electric field around the regions of most interest. The primary disadvantage of the finite element method is its computational complexity arising from grid generation, bookkeeping, and operations on large matrices. As a result, earlier investigations using this method typically have focused on small molecular systems (52). By using a compact system matrix representation and iterative equation-solving algorithms, this method has been implemented recently by You and Harvey (53) to solve large-scale macromolecular systems. But, in general, finite element algorithms are still very computationally demanding and their application in the area of molecular electrostatics has been only moderately successful.

In contrast, algorithms for solving the Poisson's or Poisson–Boltzmann equation using the finite difference method are easy to implement and readily available (54–59). In the most basic implementation, the molecule of interest is placed in a cubic box that is divided into a regular Cartesian grid at positions $r = nh$ labeled by integer vectors $n \equiv (n_x, n_y, n_z)$ and spacing $h$. The differential Eq (11) is now rewritten as a finite difference equation,

$$-\frac{1}{h^2} \sum_{n} \left( \frac{\epsilon_{n'} + \epsilon_{n}}{2} \right) (\Phi_{n'} - \Phi_n) = 4\pi \rho_n$$

(35)

where $n' = (n'_x, n'_y, n'_z)$ is summed over the 6 nearest neighboring grid points of $n$. The variables $\Phi_n$, $\rho_n$ and $\epsilon_n$ are suitably averaged over cells centered at the grid point $n$. Generally, for accurate and efficient results, one should use adaptive grids with an average spacing $h$ on the order of 1 Å or less.

In the following, we give a detailed description of a third numerical approach: the boundary element method (60–64). Unlike the finite difference method, the boundary element method is based on the concept of the reaction field rather than a direct solution to Poisson’s equation (65). Because of its mathematical rigor and sophistication, we use this method as an example to further elaborate on the essential concepts and assumptions underlying continuum electrostatic models.

3.4.2. Boundary element method. The boundary element method expresses Poisson’s equation in terms of an integral equation for the surface charge $\sigma(\vec{r})$ at the boundaries. For the cavity model, the divergence condition on the displacement vector demands that the only bound charges must be located on the molecular surface where the dielectric discontinuity occurs. Consequently the solution to Poisson’s equation is the electrostatic potential due to free charges $q_k$ and surface charges $\sigma(\vec{r})$ given earlier in Eqs (18) and (19):

$$\Phi(\vec{r}) = \frac{1}{\epsilon_p} \sum q_k \frac{1}{|\vec{r} - \vec{r}_k|} + \frac{\sigma(\vec{r})}{|\vec{r} - \vec{F}|} dS(\vec{F}).$$

(36)
As all free charges are contained in the cavity, the normal component of the electric displacement is continuous at any point on the boundary between the two dielectrics,

$$\epsilon_w \hat{n}(\mathbf{r}) \cdot E_{\text{out}}(\mathbf{r}) - \epsilon_p \hat{n}(\mathbf{r}) \cdot E_{\text{in}}(\mathbf{r}) = 0,$$

(37)

where $\hat{n}(\mathbf{r})$ is the unit normal vector on the dielectric boundary at $\mathbf{r}$. The discontinuity of the normal component of the electric field at $\mathbf{r}$ can also be expressed through the surface density of polarization charges:

$$\hat{n}(\mathbf{r}) \cdot (E_{\text{out}}(\mathbf{r}) - E_{\text{in}}(\mathbf{r})) = 4\pi \sigma(\mathbf{r}).$$

(38)

Then the surface charge density can be expressed as a function of the normal component of the electric field just outside the solute cavity:

$$\sigma(\mathbf{r}) = \frac{1}{4\pi} \frac{\epsilon_p - \epsilon_w}{\epsilon_p} \hat{n}(\mathbf{r}) \cdot E_{\text{out}}(\mathbf{r}).$$

(39)

We now derive an expression for $E_{\text{out}}(\mathbf{r}) = -\nabla \Phi_{\text{out}}(\mathbf{r})$. Since the integral in Eq (36) over the small area $(\delta A)$ around $\mathbf{r}$ is divergent, we separate the $E_{\text{out}}(\mathbf{r})$ into three terms: the field due to free charges in the cavity, the field due to surface charges outside $\delta A$, and the field due to surface charges on $\delta A$:

$$\hat{n}(\mathbf{r}) \cdot E_{\text{out}}(\mathbf{r}) = \frac{1}{\epsilon_p} \sum_k q_k \hat{n}(\mathbf{r}) \cdot (\mathbf{r} - \mathbf{r}_k) + \int_{\delta A} \frac{\sigma(\mathbf{r}') \hat{n}(\mathbf{r}) \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \, dS(\mathbf{r}') + \hat{n}(\mathbf{r}) \cdot E_{\text{local}}^{\text{out}}(\mathbf{r}),$$

(40)

where the second term is integrated over the cavity surface excluding $\delta A$. The term $E_{\text{local}}^{\text{out}}(\mathbf{r})$ refers to the contribution coming from the region, $\delta A$. To compute this, we place a closed cylinder around $\mathbf{r}$ with its bases parallel to the flat surface of $\delta A$ and walls of infinitesimally small height. In the limit of zero height walls, the solid angle corresponding to each base of the cylinder equals $2\pi$. Applying Gauss's law on this infinitesimal pillbox, we have

$$\hat{n}(\mathbf{r}) \cdot E_{\text{out}}^{\text{local}}(\mathbf{r}) = 2\pi \sigma(\mathbf{r}).$$

(41)

Collecting terms, we have the desired boundary integral equation for the surface charge,

$$\sigma(\mathbf{r}) = -\frac{\zeta}{\epsilon_p} \sum_k q_k \hat{n}(\mathbf{r}) \cdot (\mathbf{r} - \mathbf{r}_k) - \frac{\zeta}{2\pi} \int_{\delta A} \frac{\sigma(\mathbf{r}') \hat{n}(\mathbf{r}) \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \, dS(\mathbf{r}'),$$

(42)

where again $\zeta = (\epsilon_w - \epsilon_p)/(|\epsilon_w + \epsilon_p|)$.

Now we can convert this into a discrete formulation by dividing the surface of the cavity into boundary elements $A_i$, $i = 1, 2, ..., N$ centered at positions $\mathbf{r}_i$ with associated normal vectors $\hat{n}_i$ and surface charge densities $\sigma_i$. The simplest assumption is that this density is constant within each element ($61, 62$). In this case the integral equation becomes a linear algebraic equation,

$$\sigma + \sum_j K_{ij} \sigma_j = E_{ri},$$

(43)

where $K_{ij}$ represents the normalized contribution of the $j$th element to the $i$th element,
\[ K_{ij} = \xi \int_{\mathcal{A}} \frac{\hat{n}_i \cdot (\vec{r}_j - \vec{r}_i')}{2\pi|\vec{r}_j - \vec{r}_i'|^3} \, dS(\vec{r}_i'). \]  \hspace{2cm} (44)

and the inhomogeneous term comes from the normal contribution of the electric field to the surface element \( \mathcal{A} \), due to the free charges \( q_k \):

\[ E_i' = -\frac{\xi}{\epsilon_p} \sum_k \frac{q_k \hat{n}_i \cdot (\vec{r}_j - \vec{r}_k)}{2\pi|\vec{r}_j - \vec{r}_k|^3}. \]  \hspace{2cm} (45)

With the solution of the system of equations, the potential at any point is obtained by summing the Coulomb potentials from all of the charges, explicit and bound.

Compared with the finite difference method, the boundary element method offers a number of advantages \((62, 63, 65, 66)\). First and foremost, it exploits 2D elements rather than 3D grids and the elements are taken over a finite enclosed surface. The description of molecular shape can be very accurate and there is no error due to including only a finite volume of space. Furthermore, because the elements are used only on the surface of the molecule, the macromolecular charges are located on atomic centers, rather than being distributed onto the grid points of a regular three-dimensional lattice. Finally, just as in the finite element method, efficient algorithms can be easily designed to have a fine mesh in regions of particular interest and a coarser mesh where lower resolution is appropriate.

3.4.3. Visualization of electrostatic molecular surface. Structural biologists and computer scientists have long been fascinated by the remarkable complexity of the molecular surface of proteins. Considerable amount of efforts have been invested to develop rapid algorithms for the construction and rendering of molecular surfaces \((67, 68)\). With the development of fast numerical methods to solve the Poisson–Boltzmann equation for solute molecules that have complex shapes and charge distributions, new computational tools are now available that visualize the molecular surfaces of proteins with colors indicating the electrostatic potential \((69)\). This representation has become a standard tool in structural biology. Recent reports of protein structures often include color-coded pictures representing solutions to the Poisson–Boltzmann equation \((59)\).

One essential mechanism underlying most biological processes is molecular recognition, exhibited by the association of a protein and its ligand \((20)\). These associations involve interactions between accessible portions of each molecule's surface and are thought to be determined largely by the details of geometric and chemical complementarity \((70, 71)\). With regard to electrostatics, visual representation has shown that many intermolecular interactions involve associations between surfaces with complementary electrostatic potentials \((59, 72)\). Upon complex formation, the unfavorable desolvation of the interacting surfaces of the guest ligand and host protein is overcome by the formation of favorable interactions. The calculation of the electrostatic potential at the molecular surface offers a useful tool for understanding the mechanisms involved in molecular association and developing successful strategies for structure-based molecular design.

3.5. Dielectric constants

Over the past decade, very exciting scientific progress has been accomplished using the formal approach of classic electrostatic theory \((59, 73, 74)\). The misconception is
that the reliability of continuum electrostatic calculations is guaranteed. It should be pointed out that large-scale macroscopic theory was derived by considering the average electrostatic interactions for distances much larger than atomic dimensions (22). To apply this theory to detailed studies of molecular interactions, all the key electrostatic components must be carefully considered and properly calibrated. The major limitations of the continuum model are due to the physically ambiguous parameter $\epsilon$ and the structural fluctuation observed in proteins (75, 76). The heterogeneous properties of proteins and water molecules near proteins are sacrificed in the continuum model, even if the shapes of the proteins are precisely accounted for.

3.5.1. Water as dielectric continuum. Water affects the electrostatic energy in two ways: it can screen the interactions of solute charges; and it can interact directly with the charges on the macromolecule. In principle, these solvation effects can be studied using a fully realistic atomic representation of protein and solvent molecules. With this model, the simple Coulomb formula for electrostatic interactions can be incorporated into the molecular mechanics force field and applied to molecular dynamics (MD) simulations. The simulated motion would, over the long term, produce the correct solvation effects due to the average reorientation of the water molecules. But this approach requires enormous amounts of computer time and has very serious convergence problems even for the simple case of evaluating the average polarization of water molecules around an ion (22). The convergence problems become much more serious in treating proteins in solution in which the shape of the solute–solvent potential is very complicated and the number of solvent molecules needed to surround the macromolecule is very large. Thus, although direct simulation of water–protein systems can provide reliable information about the ‘structure’ of water molecules around the protein, it is necessary to find more practical and efficient alternatives for treating electrostatic solvation effects in proteins for which the energies are of primary concern.

The successful reproduction of solvation energies by various continuum models (8, 53, 65, 77) suggests that treating solvent as a dielectric continuum can provide reasonable estimates on the electrostatic energy of solvation. Interestingly, the Langevin dipole (LD) model, originated from microscopic considerations and applied to protein electrostatics almost a decade ahead of continuum numerical methods (22, 28), parallels continuum treatment in several respects. In this model, the water molecules are represented by point dipoles on a grid constructed around the protein. Each dipole is polarized towards the local field resulting from the protein atoms as well as other solvent dipoles. The polarization of a given solvent dipole in its local field is approximated by a Langevin-type function. The spacing of the grid can be chosen to reproduce the density of the actual solvent. In practice, the use of grid points with a spacing much smaller than that between real solvent molecules gives more accurate results.

3.5.2. What is the dielectric constant of proteins? The dielectric constant that should be used in calculations of protein properties depends on the part of the system which is not included explicitly in a given model (22, 74, 76). In the current macroscopic model, protein permanent dipoles are represented in the form of partial charges. There are still two factors left which contribute to the dielectric response: induced dipoles and the reorientations of charge-bearing groups. Simulations of flexible molecules need only account for the electronic polarizability, which is normally approximated by a uniform dielectric constant ($\epsilon_r$). A number of studies have determined empirically that
a choice of $\epsilon_p = 2$ or 4 for the interior of a protein provides a reasonable approximation for the effects of polarization in proteins when medium- and long-range electrostatic interactions are modeled (22, 47, 78). However, correct evaluation of the microscopic dielectric effect in molecules is very important for obtaining reliable short-range energies. In rigid systems, the local dielectric constant must also account for the reorientation effects.

In principle, it is possible to refine the uniform cavity model to include the variation of polarizability throughout a protein molecule. The dielectric properties of proteins have been investigated by a number of studies (79–83). Nakamura et al. (80) have calculated a 'local dielectric constant' from the electronic polarization of atoms and the orientational polarization of local dipoles. The latter was determined from the fluctuation–dissipation theorem with the aid of normal mode analysis. The resulting local dielectric constants ranged from 1 to 20 inside the protein. More rigorous modeling was done by Ottung and co-workers who approached the problem using quantum mechanics (79, 84) and by King et al. (81) who analyzed the protein dielectric constant from the trajectories of precise simulations, including the water molecules. The conclusions are essentially the same.

The utilization of either a uniform or a 'local' dielectric constant assumes that the response to the field is linear (45). In situations where conformational fluctuations can take place, it may not be possible to model the response in this fashion and flexibility will have to be included explicitly. The continuum approaches are based on static protein structures and thus do not adequately describe specific interactions between the protein and the nearby water molecules under such conditions (76). In principle, this problem can be solved by sufficient sampling of all relevant conformations with appropriate probabilities and averages. However, such refinement strains the limits of even our fastest computers. Serious attempts in this direction that involve the combination of a continuum model with molecular dynamics are being pursued (85–87). In this approach, the hydration forces determined by continuum electrostatics are incorporated into the molecular mechanics force field.

4. Phenomenological Studies of Electrostatic Interactions

4.1. The distribution of charged groups

At physiological pH, ionizable amino acids as well as the N- and C-termini of a protein are at least partially charged (20). In native protein structures, the groups are typically driven to the surface by solvation forces where they interact with the hydration shell or other charged groups (88). When charged groups are positioned in the interior of a protein, as they occasionally must be, oppositely charged ionizable groups are usually found within 4 Å, and the two charges form an ion pair (salt bridge) which offsets the unfavorable energy change due to removing charged moieties from water (89). When no neighboring charge of opposite sign is present, the buried charged group is usually found to be involved in a hydrogen bond(s) with a nearby polar group or groups that are electrically neutral (76). There has been no observation of isolated charged groups in the hydrophobic interiors of proteins (76, 90).

It is generally agreed that the hydrophobic effect (i.e. the favorable free energy change accompanying desolvation of nonpolar groups) is the major contributor to the
stabilization of the folded structure of globular proteins (91). However, early de novo protein design experiments which employed a substantial number of hydrophobic residues yielded proteins that had only molten globule structures (92). It is clear that the characteristic distribution of charged and polar groups plays an important role in generating unique structures and conferring conformational specificity (19).

The association of proteins involves interactions between accessible regions of the two molecules. The binding interface is formed by residues from protein surfaces, which, on average, contain more charged and polar groups (90). As a result, in contrast to the protein interior, ion-pairs are frequently observed on the interfaces of protein–protein complexes (93, 94). In general, protein active sites often provide electrostatic complementarity to the charge distribution of the binding substrate, although the compositions of the hydrophilic residues in these sites vary from case to case. Recent surveys of protein–protein complexes consistently indicate a strong positive correlation between the number of hydrophilic bridges (including both salt bridges and hydrogen bonds) and the measured binding free energy (94).

There are two classes of experimental methods that have been used to measure the contribution of charged groups to protein structure and stability: $pK_a$ shifts and site-specific mutagenesis.

4.2. $pK_a$ shift and protein as a solvent!

The ionization of each group on a protein is affected by its environment: by the protein, by the solvent, and by the ionization of other groups on the protein (21, 95). This can be detected experimentally through $pK_a$ titration of individual groups using spectroscopic methods. Following (21), the $pK_a$ shift of an acid group ($\text{AH} \rightleftharpoons \text{A}^- + \text{H}^+$) in protein can be calculated by

$$pK_a^p - pK_a^w = \frac{1}{2.3RT} \left( \Delta G_{w-p}^{\text{solv}}(\text{A}^-) - \Delta G_{w-p}^{\text{solv}}(\text{AH}) \right),$$

(46)

where $pK_a^w$ is the $pK_a$ value of the group in fully solvated state, $pK_a^p$ is the apparent $pK_a$ value of the group at its native protein environment, and $\Delta G_{w-p}^{\text{solv}}(\text{A}^-)$ and $\Delta G_{w-p}^{\text{solv}}(\text{AH})$ are, respectively, the changes in solvation energies of $\text{A}^-$ and $\text{AH}$ upon transferring from water to the corresponding protein site. A similar expression can be written for the $pK_a$ shift of a basic group ($\text{BH}^+ \rightleftharpoons \text{B} + \text{H}^+$).

It has been noted that the protonation/deprotonation of one group is influenced by the charged states of other ionizable groups in a protein (23, 96). It is useful to define an intermediate quantity, $pK_{a\text{int}}$, as the $pK_a$ value of the site when all other ionizable groups in the protein are neutral (50, 23). The $pK_{a\text{int}}$ for each site is obtained first and the mutual influence of charging one group on the titration behavior of another group is considered subsequently. Since a protein typically contain tens or hundreds of titratable sites, a complete treatment of such a multiple-site titration problem involves the examination of a large number of ionization states. To overcome the computational obstacle, various methods for treating the energetics and forces associated with ionizable groups with a minimum of computer time have been developed (96–98).

Equation (46) relates the $pK_a$ shift to the change in solvation energies associated with the transfer of two species ($\text{A}^-$ and AH) from water to a protein site. To understand
the characteristics of the local protein environments of ionizable groups, we consider an idealized spherical protein with low dielectric constant \( \epsilon_p = 2 \). Since the solvation energy of the charged compound \( (A^-) \) is much larger than that of its neutral form (AH), we concentrate on the term \( \Delta G_{w-p}^{solv}(A^-) \). The isolated acid group is approximated by a sphere of radius 2 Å; based on the Born formula, the solvation energy of this system is \(-40 \text{kcal/mol}\). The solvation energy of \( A^- \) in a nonpolar sphere of radius 20 Å depends on the distance of the charge from the center of the sphere, ranging from \(-4 \text{kcal/mol} \) (at the center) to \(-36 \text{kcal/mol} \) (on the surface). Thus, the solvation energy change \( \Delta G_{w-p}^{solv}(A^-) \) varies from 36 kcal/mol \((\sim 26 \text{ pK}_a \text{ units})\) to 4 kcal/mol \((\sim 3 \text{ pK}_a \text{ units})\). These results contradict the experimental finding that the changes in \( \text{pK}_a \) are typically very small (less than 3 pH units) and can be both positive and negative. The only explanation is that the actual microenvironment around ionized groups in proteins must be very polar \((22)\). Therefore, proteins present polar sites that can 'solvate' and 'stabilize' charged groups effectively (energetically comparable to the solvation effect of water) \((42, 99)\).

\( \text{pK}_a \) titration of the ionizable groups in proteins directly yields important information about the charge-charge interactions of the native state. With a proper reference state, \( \text{pK}_a \) shifts can be attributed to electrostatic interactions among the different ionized groups. For example, by measuring the \( \text{pK}_a \) shift of the amino group between active and inactive states, a contribution of 2.9 kcal/mol was assigned to the salt bridge between the amino terminus and the side chain of Asp 194 in the active conformation of chymotrypsin \((100)\). In T4 lysozyme the salt bridges between His-31 and Asp-70 account for the maximum in stability near pH 5 observed for this protein. Dahlquist and co-workers measured the \( \text{pK}_a \) shift of both residues in the folded structure relative to unfolded and found that the presence of Asp-70 shifts the \( \text{pK}_a \) of His-31 upward by 2.3 units at 10°C, corresponding to an interaction free energy of about 3 kcal/mol \((101)\).

4.3. Mutational study

More detailed characterization of ion-pair contributions to protein stability comes from site-specific mutagenesis studies. Both exposed and buried salt bridges have been examined \((102, 103)\). The surface salt bridge network among Asp-8, Asp-12, and Arg-110 in barnase has been analyzed by Horovitz et al. \((102)\) who mutated individual residues with different combinations to alanine (Ala). They found that the apparent contribution to the stabilization energy of the protein by a salt bridge between Asp-12 and Arg-110 is \(-1.25 \text{kcal/mol} \), whereas that of the salt bridge between Asp-8 and Arg-110 is \(-0.98 \text{kcal/mol} \). Quite opposite results have been obtained for salt bridges buried in the protein interior \((103)\). By replacing a solvent-inaccessible salt-bridge triad (Arg-31, Glu-36 and Arg-40) in Arc repressor by hydrophobic residues, Waldburger et al. found that the active mutants are more stable than the wild type by 1.4 to 4.5 kcal per mole of dimer \((103)\).

An alternative approach is to introduce 'stabilizing' salt bridges in protein. In general, such attempts have either destabilized proteins or only increased the stability by a small amount \((104, 105, 106)\). Sali et al. \((105)\) constructed the double mutant protein S28E/A32K to introduce a salt bridge and found that the interaction between Glu-28 and Lys-32 contributes only 0.2 kcal/mol to the stability of barnase. Using T4 lysozyme as a model, Dao-Pin et al. \((106)\) mutated T115E, Q123E and N144E to introduce new
salt bridges with one or more existing charged groups in the protein. X-ray diffraction data indicate that only the N144E mutants form a 2.8 Å hydrogen bond which resulted in a very small increase in the thermal stability of about 0.5 kcal/mol.

Taken together, these experimental data indicate that the net effect of charge–charge interactions is strongly context-dependent. The interaction energy of an ion pair or salt bridge is generally just enough to compensate for the unfavorable desolvation free energy of the individual charged groups. Despite the complexity, there appears to be a generalizable difference between ion pairs exposed to the solvent and those buried in the protein interior. Calculations based on numerical solutions to the Poisson–Boltzmann equation as well as analytical models indicate that a single ion pair on the protein surface generally tends to contribute more favorably to protein stability than the salt bridges buried in the protein interior (75, 76, 107). However, the net salt bridge interaction may as well depend on the interactions of the two partners with other polar and charged groups in the protein. The overall energy may become more favorable if the salt bridge also forms hydrogen bond networks with the surrounding polar or charged groups in the protein (75, 107). Thus, the composition of the microenvironment of ion pairs plays an important role in determining their overall contribution to protein stability.

This concept also explains the statistical finding that salt bridges across the binding interfaces of protein–protein complexes occur more frequently than ion-pairs located in the interior of proteins (94). The different contributions of salt bridges to folding and binding arise from the different environments to which the involved charge groups are exposed before and after the bridges are formed. These groups are more solvated in a denatured protein before folding than on the surface of the active proteins before binding. After binding, ion pairs are buried in an environment which consists of residues from the surfaces of the attending proteins, which, on average, are more hydrophilic than residues in the protein interior. As a result, the desolvation cost of an ion pair is lower, and the favorable interactions between the ion pair and its surrounding residues are generally stronger in binding than in folding.

5. Summary

In this paper, we have reviewed some fundamental concepts and methodologies in treating electrostatics in biomolecular systems. The electronic properties of proteins, the electronic polarizability of protein atoms and the orientational polarization effects of water molecules were described in rigorous microscopic terms. The continuous models that represent the protein and solvent by two different dielectric continua were then discussed. We placed the electrostatic solvation effect in the framework of the reaction field theory and emphasized the relevance of various computational treatments to the central problem of dielectric discontinuity between solvent water and solute protein. In particular, we formulated the separation of self-energies of individual charges and solvent screening of charge–charge interactions from the overall solvation effect in an attempt to clarify one of the most frequently misunderstood issues in continuum electrostatic treatments. We also covered numerical methods to solve protein systems with boundaries that better approximate the molecular surface; of particular interest is the boundary element method whose advantages potentially allow for
fast and accurate evaluation of electrostatic solvation effects in biomolecular systems. The principal error in using a dielectric continuum arises from the physically ambiguous parameter $\varepsilon$. Although electronic polarizability of proteins in a static conformation can be approximated quite well with a uniform dielectric constant, current continuum models generally fail to represent the orientational effects of charge-bearing groups and the structural fluctuations of proteins. Serious efforts in elucidating the microscopic mechanisms of electrostatic phenomena in proteins continue to provide insight and bring improvement to efficient continuum treatments.

Acknowledgements
This work was supported by NSF grant numbers DBI-9630188 (C. D., S. V., and R. C. B.) and GER-9452651 (S. R. K.), and Boston University College of Engineering (C. Z. and Z. W.). We thank B. Boghosian for helpful discussions and A. Warshel for sending us his thoughtful comments.

References


(58) R. E. Brucoleri, “Grid positioning independence and the reduction of self-energy in the


(102) A. Horovitz, L. Serrano, B. Avron, M. Bycroft, and A. R. Fersht, "Strength and co-


Appendix

Derivation of an integral form solution of the Kirkwood–Tanford model

As is well known, Eq (31) can be rewritten in a more convenient form. Let \( \rho = r r' / a^2 \) and \( z = \cos \theta \) so that

\[
\Phi_{\epsilon_s} = \frac{\epsilon_p - \epsilon_w}{\epsilon_p + \epsilon_w} \sum_{n=0}^{\infty} \frac{\rho^n a^n}{1 + \epsilon_w n + 1} P_n(z).
\]  

(47)

The numerator inside the sum can be written in terms of the generating function

\[
\sum_{n=0}^{\infty} \rho^n P_n(z) = \frac{1}{\sqrt{1 + \rho^2 - 2\rho z}}.
\]

We proceed by manipulating the factor in the summation term in Eq (47):

\[
\Phi_{\epsilon_s} = -\frac{\xi}{a} \left\{ \frac{1}{\epsilon_p + \epsilon_w} \sqrt{1 + \rho^2 - 2\rho z} \right\} \sum_{n=0}^{\infty} \frac{\rho^n}{\epsilon_p + \epsilon_w} P_n(z) + \sum_{n=0}^{\infty} \frac{\rho^n}{\epsilon_p + \epsilon_w} P_n(z)
\]

\[
= -\frac{\xi}{a} \sqrt{1 + \rho^2 - 2\rho z} \rho^\delta P_{n+\delta}(z),
\]

where \( \xi = (\epsilon_p - \epsilon_w)/(\epsilon_p + \epsilon_w) \) and \( \delta = \epsilon_w/(\epsilon_p + \epsilon_w) \).

We notice that the remaining sum can be converted to an integral whose evaluation can be carried out very rapidly. Such a conversion has apparently not been carried out in the past. It proceeds as follows.

Term by term differentiation of the sum with respect to \( \rho \) gives:

\[
\frac{d}{d\rho} \sum_{n=0}^{\infty} \rho^{n+\delta} P_n(z) = \sum_{n=0}^{\infty} \rho^{n+\delta-1} P_n(z) = \rho^{\delta-1} \frac{1}{\sqrt{1 + \rho^2 - 2\rho z}}.
\]

We can thus write the dielectric contribution to the potential in the following integral form:

\[
\Phi_{\epsilon_s} = -\frac{\xi}{a} \sqrt{1 + \rho^2 - 2\rho z} - \frac{\xi}{\epsilon_p + \epsilon_w a} \rho^{-\delta} \int_0^\rho \frac{(\rho')^{\delta-1} \rho'}{\sqrt{1 + \rho'^2 - 2\rho' z}}.
\]

The limits of integration can be rescaled so that they are independent of \( \rho \). This gives our final result (Eq (33)).