Boundary Element Solution of Macromolecular Electrostatics: Interaction Energy between Two Proteins

Huan-Xiang Zhou
National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892 USA

ABSTRACT The boundary element technique is implemented to solve for the electrostatic potential of macromolecules in an ionic solution. This technique entails solving surface integral equations that are equivalent to the Poisson and the Poisson-Boltzmann equations governing the electrostatic potential inside the macromolecules and in the solvent. A simple but robust method is described for discretizing the macromolecular surfaces in order to approximate the integral equations by linear algebraic equations. Particular attention is paid to the interaction energy between two macromolecules, and an iterative procedure is devised to make the calculation more efficient. This iterative procedure is illustrated in the electron transfer system of cytochrome *c* and cytochrome *c* peroxidase.

INTRODUCTION

We have been developing methods to study the effect of electrostatic interaction on protein-protein association kinetics. In a previous paper (Zhou, 1993) we transformed the Poisson and the Poisson-Boltzmann equations governing the electrostatic potential of two macromolecules into integral equations on their surfaces. To solve these integral equations we approximated the macromolecules as spheres (termed the two-sphere model for later references). The surface potentials were then found by spherical harmonic expansions, and the interaction energy between the molecules was obtained. However, to model macromolecular systems realistically, irregularly shaped surfaces have to be dealt with.

The most popular current method for solving the electrostatic problem of realistic macromolecules is the finite difference technique (for a review, see Davis and McCammon, 1990). Warwicker and Watson (1982) pioneered this approach, in which the Poisson and Poisson-Boltzmann equations are solved by representing the infinite three-dimensional space by a finite cubic lattice. Subsequently Klapper et al. (1986) extended this method to include the effects of salt ions in the solution. It is obviously difficult to use this method to calculate the electrostatic interaction energy between two macromolecules, as two molecules separated by a large distance would involve a large-size lattice.

A newer approach to the macromolecular electrostatic problem is the boundary element technique. It was first used by Zauhar and Morgan (1985, 1988) in the case where no salt ions were present. Extensions including salt ions were recently made by two independent groups (Yoon and Lenhoff, 1990; Juffer et al., 1991). In this approach the integral equations are transformed into linear algebraic equations by discretizing the macromolecular surface into boundary elements. In the original discretization method described by Zauhar and Morgan and employed both by Yoon and Lenhoff and by Juffer et al., the intersections with the surface by spokes sent out from a fixed point inside the molecule were taken as nodes. Neighboring nodes were connected to form a web of small triangles. This web then approximated the macromolecular surface. As noted by Zauhar and Morgan (1988), this scheme is successful only if the distribution of nodes is not too irregular. A more recent discretization method by Zauhar and Morgan (1990) remedies this failure. However, it appears to be very complicated.

So far the boundary element method has been formulated only for a single macromolecule in an ionic solution. Even though a recent application by Yoon and Lenhoff (1992) dealt with the interaction between a protein and a charged wall, it essentially reduces to the problem of an isolated protein. In this paper we describe a simple but robust method for discretizing macromolecular surfaces and implement the boundary element technique for an arbitrary number of macromolecules in an ionic solution. The boundaries chosen to be discretized are the solvent-accessible surfaces, i.e., the surfaces swept by the center of a solvent atom when it is rolled over the macromolecules. Because of our own interest in the protein-protein association process, particular attention is paid to the electrostatic interaction energy between two proteins. An iterative procedure is specifically devised to make its calculation efficient.

The rest of the paper is organized as follows. We first present the integral equations for an arbitrary number of macromolecules in an ionic solution. Next, our method of discretizing their solvent-accessible surfaces is described. This method is then applied to two electron-transfer proteins, cytochrome *c* and cytochrome *c* peroxidase, when each is isolated in an ionic solution. For each protein the surface potential and its normal derivative are found by a matrix inversion, from which the potential at any point can be obtained. Several tests are performed on the two proteins to ensure the reliability and accuracy of the discretization method.

After these preparations we turn our attention to the main focus of the paper, the interaction energy between two macromolecules. From the general linear algebraic

Received for publication 8 March 1993 and in final form 29 April 1993.
Address reprint requests to Dr. Huan-Xiang Zhou, Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, Building 5, Room 136, National Institutes of Health, Bethesda, MD 20892.

0006-3495/93/08/955/09 $2.00

equations approximating the integral equations for the electrostatic potential of the two molecules, we derive iterative algebraic equations for the changes in electrostatic potential due to the presence of the second molecule. Compared to direct matrix inversion, iteration results in the saving of an enormous amount of computational time. This crucial step enables the calculation of the interaction energy between the molecules in many configurations. The iteration procedure is first checked against the previous study of the two-sphere model and then applied to the electron transfer proteins cytochrome *c* and cytochrome *c* peroxidase in a number of different configurations. The final part of the paper discusses possible applications of the present work.

INTEGRAL EQUATIONS

In this section we present the integral equations governing the electrostatic potential of a system with an arbitrary number (M) of macromolecules in an ionic solution. Each macromolecule is separated from the solvent by its solvent-accessible surface. Inside the ith surface S_i, the electrostatic potential $V_i = V(\mathbf{r}_i)$ satisfies the Poisson equation. An elementary derivation via Green's theorem (Jackson, 1962) shows that it is equivalent to the integral equation

$$4\pi V_i = \frac{4\pi v_i}{\epsilon_{\text{in}}} + \int_{S_i} dS' \left(g' \frac{1}{\epsilon_{\text{in}} |\mathbf{r}_i - \mathbf{r}_{s'}|} - f' \mathbf{n}' \cdot \nabla_{s'} \frac{1}{|\mathbf{r}_i - \mathbf{r}_{s'}|} \right), \tag{1a}$$

$$f' = V(\mathbf{r}_{s'}), \qquad g' = \epsilon_{\text{in}} \mathbf{n}' \cdot \nabla_{s'} V(\mathbf{r}_{s'}). \tag{1b, c}$$

In the above v_i is the vacuum potential at $\mathbf{r}_i$ produced by the charges of the ith molecule, $\mathbf{n}'$ is the unit vector pointing outward along the normal of the surface S_i at $\mathbf{r}_{s'}$, and ϵ_{in} is the interior dielectric constant. When the position $\mathbf{r}_i$ is moved onto the surface S_i ($\mathbf{r}_i \to \mathbf{r}_{si}$) we have the following self-consistent relation for the surface potential f_i and its normal derivative g_i,

$$2\pi f_i = \frac{4\pi v_{si}}{\epsilon_{\text{in}}} + \int_{S_i} dS' \left(g' \frac{1}{\epsilon_{\text{in}} |\mathbf{r}_{si} - \mathbf{r}_{s'}|} - f' \mathbf{n}' \cdot \nabla_{s'} \frac{1}{|\mathbf{r}_{si} - \mathbf{r}_{s'}|} \right). \tag{2}$$

Similarly, the Poisson-Boltzmann equation satisfied by the electrostatic potential $V_o = V(\mathbf{r}_o)$ in the solvent is equivalent to the integral equation

$$-4\pi V_o = \sum_{j=1}^{M} \int_{S_j} dS' \left(g' \frac{e^{-|\mathbf{r}_o - \mathbf{r}_{s'}|/\lambda}}{\epsilon_{\text{ou}} |\mathbf{r}_o - \mathbf{r}_{s'}|} - f' \mathbf{n}' \cdot \nabla_{s'} \frac{e^{-|\mathbf{r}_o - \mathbf{r}_{s'}|/\lambda}}{|\mathbf{r}_o - \mathbf{r}_{s'}|} \right), \tag{3}$$

where ϵ_{ou} is the outside dielectric constant and λ is the Debye-Huckel screening length. The boundary conditions for the electrostatic potential have been built into Eq. 3. When the position $\mathbf{r}_o$ is moved onto the ith surface $S_i(\mathbf{r}_o \to \mathbf{r}_{si})$, we have another self-consistent relation for f_i and g_i,

$$-2\pi f_i = \sum_{j=1}^{M} \int_{S_j} dS' \left(g' \frac{e^{-|\mathbf{r}_{si} - \mathbf{r}_{s'}|/\lambda}}{\epsilon_{\text{ou}} |\mathbf{r}_{si} - \mathbf{r}_{s'}|} - f' \mathbf{n}' \cdot \nabla_{s'} \frac{e^{-|\mathbf{r}_{si} - \mathbf{r}_{s'}|/\lambda}}{|\mathbf{r}_{si} - \mathbf{r}_{s'}|} \right). \tag{4}$$

Eqs. 2 and 4 determine the surface potentials f_i and their normal derivatives g_i on the solvent-accessible surfaces. Once they are determined, Eq. 1a can be used to determine the potentials in the interiors and Eq. 3 can be used to determine the potential in the outside.

DISCRETIZATION OF SOLVENT-ACCESSIBLE SURFACES

In order to solve the integral equations of the last section, Eqs. 2 and 4, we have to discretize irregularly shaped solvent-accessible surfaces, that is, to tile the surfaces with small boundary elements (BEs). This discretization process is the most complicated part of the boundary element technique. Through discretization the integral equations are transformed into linear algebraic equations. Subsequently the surface potentials and their normal derivatives at the BE centers can be found by a simple matrix inversion.

A useful discretization scheme has to work for any surface. We now describe a simple method that achieves this goal. It discretize the solvent-accessible surface with uniformity (i.e., the BEs have roughly the same area) and distinguish cavity and channel surfaces from the exterior surface (cavity and channel surfaces may be discarded from the solvent-accessible surface).

As the solvent-accessible surface of a macromolecule is made of pieces of spherical regions, we choose spherical rectangles as the BEs. Each BE is described by the Cartesian coordinates (x_n, y_n, z_n) of the center $\mathbf{s}_n$ of the sphere it is on, the spherical coordinates (r_n, ξ_n, ϕ_n) of the BE center $\mathbf{b}_n$ (ξ: cosine of the polar angle), and its extensions $\Delta\xi_n$ and $\Delta\phi_n$ in the polar and azimuthal angles (see Fig. 1). They are found through the following steps:

(*a*) Equally separated parallel planes, termed z-planes (separation: Δz), are used to cut through the solvent-accessible surface. At each cut, arcs making up the cross section of the solvent-accessible surface are found and sorted into one or more contiguous lists. Each contiguous list of arcs makes a closed contour termed loop.

(*b*) Each loop is checked to see whether it belongs to cavity and channel surfaces or to the exterior surface. Cavity and channel loops are the ones inside the exterior loops. For the former group, when the individual arcs (more rigorously, the circles they are on) are traced counterclockwise, the whole loop is traced clockwise. For the latter group, when the individual arcs are traced counterclockwise, the whole loop is also traced counterclockwise.

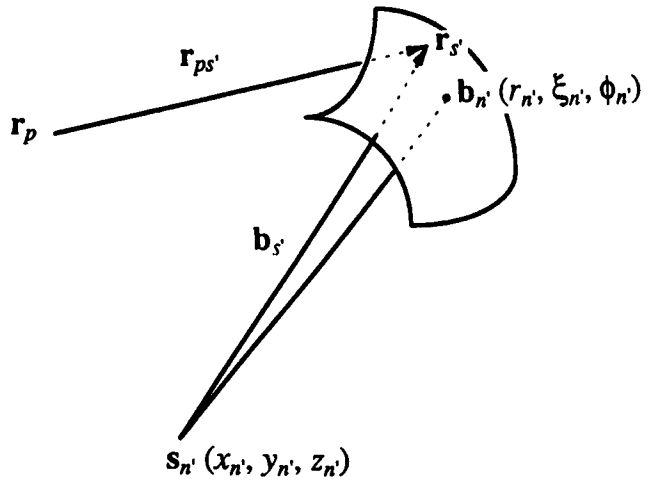


FIGURE 1 The definition of a BE.

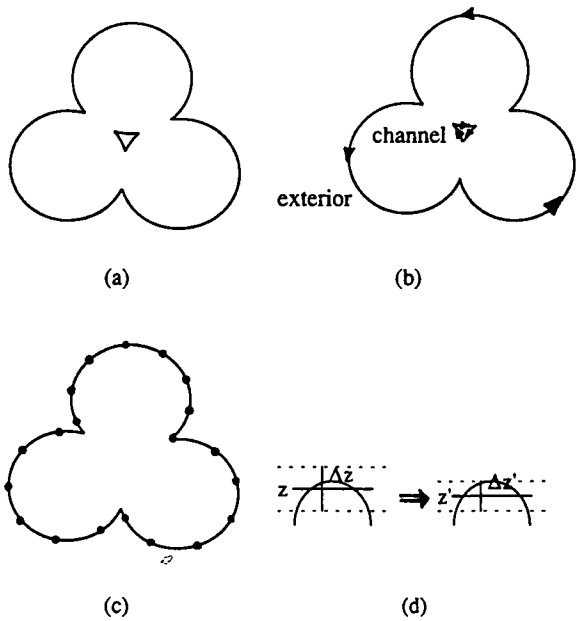


FIGURE 2 Steps in discretizing the surface of a molecule composed of three atoms located on the vertices of an equilateral triangle. (*a*) Cutting through the molecule with z-planes and sorting exposed arcs into loops. (*b*) Distinguishing exterior loops and cavity and channel loops. (*c*) Tracing exterior loops to select BE centers. (*d*) Adjusting the BE center polar angle and its polar angle extension when a BE is near the top of a sphere.

Based on this fact a simple algorithm can be devised to distinguish the two groups. The details are given in the Appendix. After this distinction the cavity and channel loops may be discarded.

(*c*) Each remaining loop is traced and at equal intervals (length: Δl), points are taken as BE centers. By keeping track of which sphere and which arc each BE center is on, (x_n, y_n, z_n) and (r_n, ξ_n, ϕ_n) can be found. As for the BE extensions, Δz divided by the sphere radii r_n gives $\Delta\xi_n$, while Δl divided by the arc radii gives $\Delta\phi_n$.

(*d*) An important exception in step (*c*) occurs when a particular z-plane intersects a spherical part of the solvent-accessible surface, but only one of its neighboring z-planes intersects that sphere. This happens when that particular z-plane is near the top or bottom of the sphere. In this case the BE centers originally on the particular z-plane are adjusted so that their polar angle extensions $\Delta\xi_n$ extend or shrink to the top or bottom of the sphere and ξ_n are moved to the centers of the polar angle extensions. All other parameters remain the same.

These four steps are illustrated in Fig. 2 for the case of a molecule composed of three atoms located on the vertices of an equilateral triangle.

Having discretized the solvent-accessible surfaces in the above manner, we now can transform the integral equations for the surface potentials f_i and their normal derivatives g_i into linear algebraic equations. We assume that f_i and g_i are uniform over each BE; thus the surface integrals over the solvent-accessible surfaces reduce to sums over the BEs. Eqs. 2 and 4 become

$$\frac{1}{2}\mathbf{f} = \frac{1}{\epsilon_{in}}\mathbf{v} + \mathbf{A}\cdot\mathbf{f} + \frac{1}{\epsilon_{in}}\mathbf{C}\cdot\mathbf{g}, \tag{5}$$

$$-\frac{1}{2}\mathbf{f} = \mathbf{B}\cdot\mathbf{f} + \frac{1}{\epsilon_{ou}}\mathbf{D}\cdot\mathbf{g}. \tag{6}$$

Specifically for the case of two macromolecules, the vectors **f**, **g**, and **v** have the form

$$\mathbf{f} = \begin{pmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{pmatrix}, \tag{7}$$

the matrices **A** and **C** have the form

$$\mathbf{A} = \begin{bmatrix}\mathbf{A}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{A}_2\end{bmatrix}, \tag{8a}$$

and the matrices **B** and **D** have the form

$$\mathbf{B} = \begin{bmatrix}\mathbf{B}_1 & \mathbf{B}_1'\\ \mathbf{B}_2' & \mathbf{B}_2\end{bmatrix}. \tag{8b}$$

Generalization to a system with an arbitrary number of macromolecules is straightforward.

The vector elements $\mathbf{f}_i$, $\mathbf{g}_i$, and $\mathbf{v}_i$ are, respectively, the surface potential, its normal derivative, and the vacuum potential due to the charges of the ith molecule on the BEs of the ith molecule. The matrix elements $\mathbf{A}_i$, $\mathbf{C}_i$, $\mathbf{B}_i$, $\mathbf{B}_i'$, $\mathbf{D}_i$, and $\mathbf{D}i'$ are surface integrals involving the Green functions over the BEs. They all can be expressed by one of the following two general forms (cf. Fig. 1):

$$I_1(\mathbf{r}_p, BE_{n'}, \kappa) = \frac{r_{n'}}{4\pi}\int_{\phi_{n'}-\Delta\phi_{n'}/2}^{\phi_{n'}+\Delta\phi_{n'}/2} d\phi_{s'}\int_{\xi_{n'}-\Delta\xi_{n'}/2}^{\xi_{n'}+\Delta\xi_{n'}/2} d\xi_{s'}\,\frac{\mathbf{r}_{ps'}\cdot\mathbf{b}_{s'}}{r_{ps'}^3} \times (1+\kappa r_{ps'})e^{-\kappa r_{ps'}}, \tag{9a}$$

$$I_2(\mathbf{r}_p, BE_{n'}, \kappa) = \frac{r_{n'}^2}{4\pi} \int_{\phi_{n'}-\Delta\phi_{n'}/2}^{\phi_{n'}+\Delta\phi_{n'}/2} d\phi_{s'} \int_{\xi_{n'}-\Delta\xi_{n'}/2}^{\xi_{n'}+\Delta\xi_{n'}/2} d\xi_{s'} \frac{1}{r_{ps'}} e^{-\kappa r_{ps'}}, \quad (9b)$$

where $BE_{n'}$ denotes the parameters of a BE. Explicitly,

$$(\mathbf{A}_i)_{m_i n_i} = I_1(\mathbf{r}_{m_i}, BE_{n_i}, 0), \quad (10a)$$

$$(\mathbf{C}_i)_{m_i n_i} = I_2(\mathbf{r}_{m_i}, BE_{n_i}, 0), \quad (10b)$$

$$(\mathbf{B}_i)_{m_i n_i} = I_1(\mathbf{r}_{m_i}, BE_{n_i}, 1/\lambda), \quad (10c)$$

$$(\mathbf{D}_i)_{m_i n_i} = I_2(\mathbf{r}_{m_i}, BE_{n_i}, 1/\lambda), \quad (10d)$$

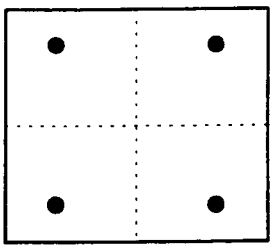
and

$$(\mathbf{B}_i')_{m_i n_j} = I_1(\mathbf{r}_{m_i}, BE_{n_j}, 1/\lambda), \quad (10e)$$

$$(\mathbf{D}_i')_{m_i n_j} = I_2(\mathbf{r}_{m_i}, BE_{n_j}, 1/\lambda). \quad (10f)$$

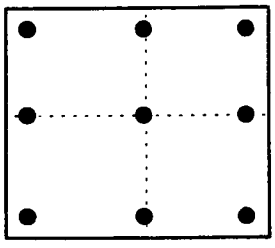
Notice that in the last two equations, the indices m_i refer to one surface whereas n_j refer to the other.

We calculate the integrals of Eqs. 9 numerically. For nondiagonal elements (i.e., $m_i \neq n_i$) a four-point weighted sum (Abramowitz and Stegun, 1964) shown in Fig. 3 *a* is used to approximate the integral over each BE. To calculate the diagonal elements, each BE is first divided into 8 × 8 subelements. The integral over each subelement is then approximated by a nine-point weighted sum (Abramowitz and Stegun, 1964) shown in Fig. 3 *b*.



(a)

(x_i, y_i)	w_i
$(\pm\sqrt{\frac{1}{3}}a, \pm\sqrt{\frac{1}{3}}b)$	1/4



(b)

(x_i, y_i)	w_i
(0, 0)	16/81
$(\pm\sqrt{\frac{3}{5}}a, \pm\sqrt{\frac{3}{5}}b)$	25/324
$(0, \pm\sqrt{\frac{3}{5}}b)$	10/81
$(\pm\sqrt{\frac{3}{5}}a, 0)$	10/81

FIGURE 3 (*a*) The scheme used in calculating the surface integral over a BE for a nondiagonal matrix element. (*b*) The scheme used in calculating the surface integral over a subelement for a diagonal matrix element. In each case a two-dimensional integral of the type $\int_{-a}^{a} dx \int_{-b}^{b} dy f(x, y)$ is approximated by a weighted sum $4ab \sum_i w_i f(x_i, y_i)$.

After calculations of the matrix elements, the surface potentials **f** and their normal derivatives **g** on the BEs can be found by a matrix inversion,

$$\begin{pmatrix} \epsilon_{ou}\mathbf{f} \\ \mathbf{g} \end{pmatrix} = \mathbf{H}^{-1} \cdot \begin{pmatrix} \mathbf{v} \\ \mathbf{0} \end{pmatrix}, \quad (11)$$

where we have defined

$$\mathbf{H} = \begin{bmatrix} (\epsilon_{in}/\epsilon_{ou})(\tfrac{1}{2}\mathbf{1} - \mathbf{A}) & -\mathbf{C} \\ \tfrac{1}{2}\mathbf{1} + \mathbf{B} & \mathbf{D} \end{bmatrix}. \quad (12)$$

We do the inversion by using *LU* decomposition (Press et al., 1986). This is the most time-consuming part of the calculation. Notice that the matrix **H** depends only on the solvent-accessible surfaces, so the same inverse can be used for different charge distributions of the macromolecules that have the same surfaces. Once **f** and **g** are found, the potential at any point can be obtained by simple summations (cf. Eqs. 1a and 3):

$$V_i = \frac{v_i}{\epsilon_{in}} + \sum_{n_i} [I_1(\mathbf{r}_i, BE_{n_i}, 0) f_{in_i} + I_2(\mathbf{r}_i, BE_{n_i}, 0) g_{in_i}], \quad (13)$$

$$-V_o = \sum_{i=1}^{M} \sum_{n_i} [I_1(\mathbf{r}_o, BE_{n_i}, 1/\lambda) f_{in_i} + I_2(\mathbf{r}_o, BE_{n_i}, 1/\lambda) g_{in_i}]. \quad (14)$$

This ends the description of our method for solving the electrostatic potential of macromolecules in an ionic solution. Next we test and apply this method in a number of illustrative cases.

TESTS ON SINGLE PROTEINS

To check the reliability and accuracy of our method, we first used it to calculate the electrostatic potentials of single proteins. An additional reason for doing these calculations is that the results can be used to find the interaction energy between two proteins, as will be seen shortly.

The first application is on the protein yeast iso-1-cytochrome *c*. For the sake of specificity the quantities of this protein will be referred to with an index "1." Its crystal structure was solved to a resolution of 1.23 Å by Louie and Brayer (1990) (PDB entry 1ycc). It has 108 amino acids and a heme and contains $L_1 = 893$ heavy atoms. We used the following atomic radii: C, 2 Å; N, 1.7 Å; O, 1.5 Å; S, 1.8 Å; and Fe, 1.7 Å. A solvent radius of 1.4 Å was used for generating the solvent-accessible surface. The partial charges on the amino acid atoms were taken from the work of McCammon et al. (1979), and those on the heme atoms were taken from the work of Northrup et al. (1981). To account for the trimethylation of residue Lys^{72}, the charge on the NZ atom was

reduced by one unit, and the three methyl groups had zero charges. The net charge on cytochrome *c* is +5e. For the dielectric constants, we chose $\epsilon_{in} = 4$ and $\epsilon_{ou} = 78.5$. Throughout this study the ionic strength is kept at 0.16 m.

Three independent calculations were performed on cytochrome *c*. For the first two calculations *z*-planes perpendicular to the *z*-axis of the X-ray structure were taken to intersect the protein. One used $\Delta z = 1$ Å and $\Delta l = 2$ Å to generate 2360 BEs, the other used $\Delta z = 2$ Å and $\Delta l = 4$ Å to generate 594 BEs. For the third calculation *z*-planes parallel to the heme plane were used, resulting in 585 BEs with $\Delta z = 2$ Å and $\Delta l = 4$ Å. The calculations of the surface potential $\mathbf{f}_1$ and its normal derivative $\mathbf{g}_1$ with the coarser discretizations took about 6 min on a Convex C240 computer, and with the finer discretization it took about 7 h. This significant time difference reflects the fact that matrix inversion is an N^3 process (N, number of BEs). The differences in the resulting potential at the atomic centers from the three calculations are shown in Fig. 4 *a*, in comparison with the potential itself (with the vacuum contribution v/ϵ_{in} subtracted). Among all the atoms the maximum difference is 2 kcal/mol/e, occurring at the O atom of Ser^{47}. In comparison $V - v/\epsilon_{in} = -14$ kcal/mol/e at this site. From the close agreement among the three independent calculations, we conclude that discretizations with $\Delta z = 2$ Å and $\Delta l = 4$ Å are sufficient for accurate calculations of the electrostatic potential of proteins.

The second application is on the electron-transfer partner of yeast cytochrome *c*, yeast cytochrome *c* peroxidase. The quantities of this protein will be referred to with an index "2." We took the 2.2-Å X-ray structure of Wang et al. (1990) for a plasmid-encoded form of cytochrome *c* peroxidase expressed in *Escherichia coli* (PDB entry 1ccp). Three amino acids at the NH_2 terminal and a number of side-chain atoms were missing in the crystal structure because of excessive motional disorders. The result is a total of 293 amino acids and a heme with $L_2 = 2382$ heavy atoms. We did not make any structural additions to the X-ray structure. The charges on the outermost side-chain atoms of Glu and Lys residues with missing atoms were adjusted so that Glu residues had a total charge of −e and Lys residues had a total charge of +e. A unit charge was also added to the heme Fe atom to reflect its ferric state. The net charge on cytochrome *c* peroxidase is −13e.

We discretized the solvent-accessible surface of cytochrome *c* peroxidase with *z*-planes both perpendicular to the X-ray structure *z*-axis and parallel to the heme plane. With $\Delta z = 2$ Å and $\Delta l = 4$ Å the two discretizations resulted in 1134 and 1133 BEs, respectively. It took about 50 min on a Convex C240 computer to calculate the surface potential $\mathbf{f}_2$ and its normal derivative $\mathbf{g}_2$ for each discretization. The difference in the resulting potential at the atomic centers from the two calculations is shown in Fig. 4 *b*, in comparison with the potential itself. Among all the atoms the maximum difference is 5 kcal/mol/e, occurring at the O atom of Pro^{145}. In comparison $V - v/\epsilon_{in} = 56$ kcal/mol/e at this site.

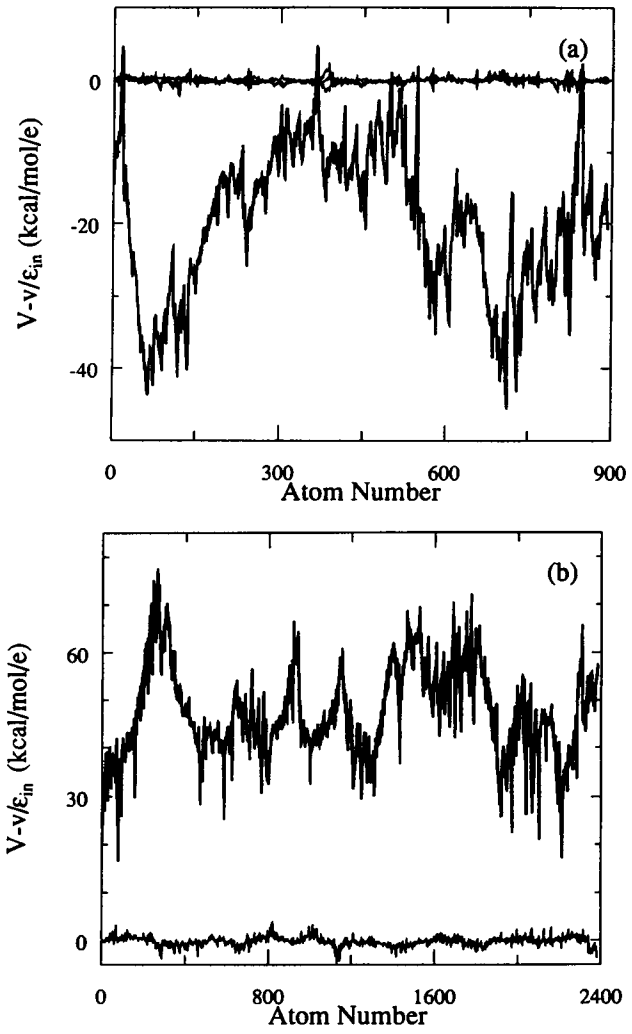


FIGURE 4 (*a*) The differences (*three thin curves at the top*) of the potential at the atomic centers of cytochrome *c* from the calculations with three different discretizations. The thick curve is the potential itself from the calculation with 2360 BEs. (*b*) The difference (*thin curve at the bottom*) of the potential at the atomic centers of cytochrome *c* peroxidase from the calculations with two different discretizations. The thick curve is the potential itself from the calculation with 1134 BEs.

INTERACTION ENERGY BETWEEN TWO PROTEINS

We now turn to the main concern of this paper, calculating the interaction energy between two proteins in an ionic solution. It is highly desirable to find an efficient method for doing this calculation if one intends to explore many different configurations of the proteins. If we were to continue what was just done for single proteins and directly invert matrices, the calculation for each configuration of the electron transfer system of cytochrome *c* and cytochrome *c* peroxidase would take about 3 h on a Convex C240 computer. To do the calculation for many configurations using direct matrix inversion would be quite impractical. For this reason an alternative method has to be found.

Our solution is to use iteration to find the changes in the surface potentials $\mathbf{f}_i$ and their normal derivatives $\mathbf{g}_i$ due to the presence of a second protein. Let these changes be $\Delta\mathbf{f}_i$ and $\Delta\mathbf{g}_i$. Comparing the linear algebraic equations satisfied by $\mathbf{f}_i$ and $\mathbf{g}_i$ and by $\mathbf{f}_i + \Delta\mathbf{f}_i$ and $\mathbf{g}_i + \Delta\mathbf{g}_i$ (Eqs. 5 and 6), we obtain for the first protein

$$\frac{1}{2}\Delta\mathbf{f}_1 = \mathbf{A}_1 \cdot \Delta\mathbf{f}_1 + \frac{1}{\epsilon_{in}}\mathbf{C}_1 \cdot \Delta\mathbf{g}_1, \tag{15}$$

$$-\frac{1}{2}\Delta\mathbf{f}_1 = \mathbf{B}_1 \cdot \Delta\mathbf{f}_1 + \frac{1}{\epsilon_{ou}}\mathbf{D}_1 \cdot \Delta\mathbf{g}_1 + \mathbf{B}_1' \cdot (\mathbf{f}_2 + \Delta\mathbf{f}_2) + \frac{1}{\epsilon_{ou}}\mathbf{D}_1' \cdot (\mathbf{g}_2 + \Delta\mathbf{g}_2), \tag{16}$$

and a similar result for the second protein. Simple manipulations lead to

$$\begin{pmatrix} \epsilon_{ou}\Delta\mathbf{f}_1 \\ \Delta\mathbf{g}_1 \end{pmatrix} = \mathbf{H}_1^{-1} \cdot \begin{pmatrix} \mathbf{0} \\ \mathbf{h}_1 \end{pmatrix}, \tag{17}$$

where

$$\mathbf{h}_1 = -[\mathbf{B}_1' \cdot \epsilon_{ou}(\mathbf{f}_2 + \Delta\mathbf{f}_2) + \mathbf{D}_1' \cdot (\mathbf{g}_2 + \Delta\mathbf{g}_2)], \tag{18}$$

and a similar result is obtained for the second protein. Notice that the matrix inversions in Eq. 17 and, similarly, for the second protein are those required for calculating the surface potentials and their normal derivatives of single proteins. If we start with $\Delta\mathbf{f}_i = 0$ and $\Delta\mathbf{g}_i = 0$ and use Eqs. 17 and 18 to iterate, then inversion of the full matrix of the two proteins can be avoided. More importantly, as all the information about the relative configuration of the two proteins shows up only in the matrices $\mathbf{B}_i'$ and $\mathbf{D}_i'$, if we invert single protein matrices once, their inverses can be used to calculate the changes in the potential and its normal derivative for any configuration of the two proteins. As each iteration is an N^2 process, compared to the N^3 process of matrix inversion, an enormous saving in computation time is expected.

After $\Delta\mathbf{f}_i$ and $\Delta\mathbf{g}_i$ are found, the changes ΔV_{il_i} in the electrostatic potentials at all the atomic centers $\mathbf{r}_{il_i}$ can be obtained by simple summations (cf. Eq. 13),

$$\Delta V_{il_i} = \sum_{n_i} [I_1(\mathbf{r}_{il_i}, BE_{n_i}, 0)\Delta f_{in_i} + I_2(\mathbf{r}_{il_i}, BE_{n_i}, 0)\Delta g_{in_i}]. \tag{19}$$

The interaction energy between the two proteins is given by

$$U = \frac{1}{2}\sum_{i=1}^{2}\sum_{l_i=1}^{L_i} q_{il_i}\Delta V_{il_i}, \tag{20}$$

where q_{il_i} are the partial charges of atoms l_i of the ith molecule. We emphasize that one has to solve the two-protein problem to obtain the interaction energy. It is incorrect to find the interaction energy through simply multiplying the electrostatic potential of one protein (when it is alone) by the charges of the other protein, as has been done in some studies (e.g., Northrup et al., 1987). The discrepancies caused by this crude treatment are shown in our earlier work (Zhou, 1993) for the two-sphere model of protein-protein association.

We first tested this iterative procedure on the two-sphere model. In this model each sphere bears three charges, one central and the other two on opposite sides of the center but separated by an equal distance from it. The noncentral charges have the same magnitude but opposite signs. For the first (second) sphere, the radius is 14 Å (21 Å), the central charge is +8e (−12e), and the noncentral charges have a magnitude of 2.29e (2.21e) and a distance of 6.2 Å (9.3 Å) from the center. Each spherical surface was discretized by dividing both its azimuthal angle and the cosine of its polar angle into 20 equal parts, resulting in 400 BEs. Ten iterations produced interaction energy that is in agreement with that obtained from direct matrix inversions to five significant digits. Whereas calculating the interaction energy for each configuration using direct matrix inversions took about 18 min on a Convex C240 computer, it took about 12 s using the iterative procedure (not including the time for inverting single sphere matrices). A reduction of 90-fold in computation time is achieved through the iterative procedure. The resulting interaction energy between the two charge-bearing spheres when they have their axes aligned is plotted as a function of their distance in Fig. 5. It is in agreement with the result of the previous study obtained by expanding the surface potentials and their normal derivatives in spherical harmonics (Fig. 6 in Zhou, 1993). Thus the previous work on the two-sphere model provides a valuable check on our new method.

After the test on the two-sphere model, we applied the iterative procedure to the electron transfer system of cytochrome *c* and cytochrome *c* peroxidase. Configurations with

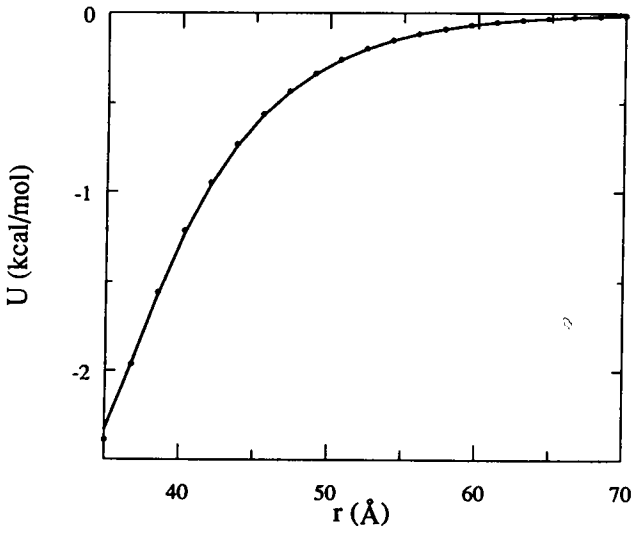


FIGURE 5 Comparison of the interaction energy of the two-sphere model as a function of their separation from the boundary element technique (*circles*) and from spherical harmonic expansions (*solid curve*).

parallel hemes and Fe-Fe distance less than 30 Å were searched to select the ones in which the proteins had their solvent-accessible surfaces in separation. Each configuration is described by a Fe-Fe distance and three orientational angles. At Fe-Fe distances of 30 Å, 29 Å, . . . , all orientational angles were sampled 20 times. Each configuration was first screened by testing whether the surface atoms of one protein, as found in the surface discretization procedure, were in contact with any surface atom of the other protein. The ones in which the surface atoms of one protein were not in contact with those of the other protein were then tested further by using all the atoms in the proteins. Of 8000 configurations at a Fe-Fe distance of 30 Å, 45 were selected. At 29 Å, six configurations were acceptable. Below 29 Å, none were found. One selected configuration is shown in Fig. 6. We calculated the interaction energy for the 51 acceptable configurations. After the inversions of single protein matrices, each configuration took about 2 min on a Convex C240 computer. When matrix and vector operations were performed via Convex intrinsic routines and the calculations were done on a Convex C3830 computer, the 2-min time was reduced to 6 s.

The energy for 10 of the acceptable configurations is shown in Fig. 7. In this figure cytochrome *c* is represented by a bond drawn from its heme Fe atom to its heme CBC atom. This bond is a part of the electron transfer path. This figure clearly shows the steering effect of the electrostatic interaction. When the Fe-CBC bond of cytochrome *c* is pointing toward the peroxidase heme, the system has a low energy. As soon as the Fe-CBC bond of cytochrome *c* points away from the peroxidase heme, the system has a high energy and thus becomes unstable. This shows that cytochrome *c* and its peroxidase have "designed" their charge distributions in a way to make electron transfer between them highly efficient.

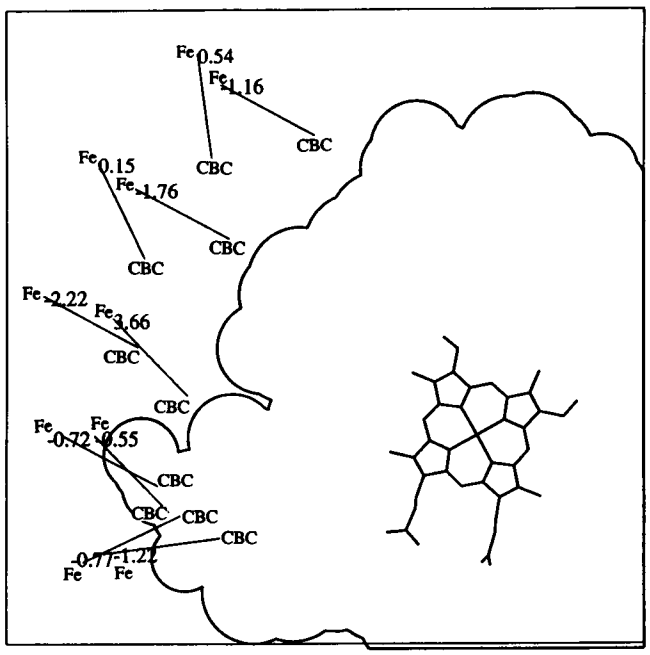


FIGURE 7 The interaction energy for ten representative acceptable configurations of the cytochrome *c* and cytochrome *c* peroxidase system. The energy is in units of kcal/mol.

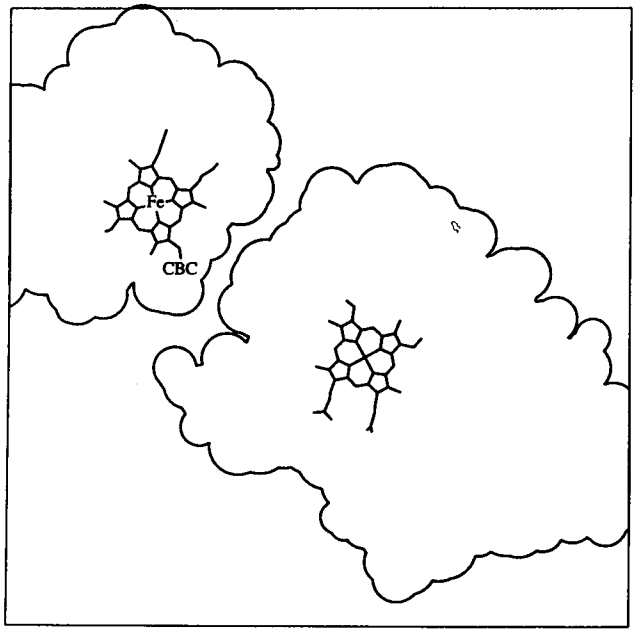


FIGURE 6 One acceptable configuration of the cytochrome *c* (*top*) and cytochrome *c* peroxidase (*bottom*) system. The heme Fe and CBC atoms of cytochrome *c* are explicitly labeled.

DISCUSSION

We have presented a method for calculating the electrostatic potential of macromolecules in an ionic solution using the boundary element technique. A general formulation involving matrix inversions is given for an arbitrary number of macromolecules. Its specialization to the case of single proteins is thoroughly tested. To calculate the interaction energy between two macromolecules efficiently, an iterative procedure is devised. Compared to direct matrix inversion, the iterative procedure is shown to reduce computation time enormously.

Previously for the two-sphere model (Zhou, 1993), by spherical harmonic expansions we were able to calculate the forces and torques on the spheres for 8×10^4 configurations in 10 h on a Convex C240 computer. Utilizing these precalculated forces and torques in Brownian dynamics simulations, we obtained the diffusional association rate of the model. It was found that the association rate is approximately proportional to the Boltzmann factor averaged over the configurations of the reaction complex. The method developed here for calculating the protein-protein

interaction energy can immediately be used to evaluate the average Boltzmann factor. A quick estimate of their diffusional association rate is then achieved by the product of this Boltzmann factor and the rate in the absence of the interaction, a quantity readily obtainable in a Brownian dynamics simulation. In addition, it appears possible to use this method for the force and torque calculations required in a full Brownian dynamics simulation of two associating proteins to directly obtain their association rate. For example, for the cytochrome *c* and cytochrome *c* peroxidase system each configuration takes 6 s on a Convex C3830 computer. At this speed the calculations for 8×10^4 configurations can be accomplished in 135 h of computer time. For the present method to become routinely applicable in Brownian dynamics simulations of protein-protein association processes, it has to be improved to further reduce the computational time.

Even though the main focus of this paper has been on the electrostatic interaction energy between two macromolecules, we believe that the developments advanced here will benefit the calculation of the electrostatics of single macromolecules. The surface discretization method presented appears to be very powerful. It is simple yet works for any macromolecular surface. For cytochrome *c* peroxidase, which has 293 amino acids and a heme, 1130 boundary elements generated via this method are sufficient to give accurate solution of the electrostatic potential. In comparison, over 4000 nodes generated via a very sophisticated method by Zauhar and Morgan (1990) were needed to calculate the potential of ferredoxin, a protein with only 54 amino acids. At least a part of this difference arises from the fact that different types of surfaces are discretized: our surface is a solvent radius (1.4 Å) away from that of Zauhar and Morgan. An interesting prospect is that by partitioning a large macromolecule into two parts and adapting the iterative procedure devised here for a two-protein system, the calculation of the macromolecular electrostatics can be made even more efficient.

An X-ray structure for the cytochrome *c* and cytochrome *c* peroxidase complex has recently appeared (Pelletier and Kraut, 1992). As the coordinates of the complex become available it will be very interesting to apply our method to it, calculate its interaction energy, and explore neighboring configurations.

APPENDIX

In this Appendix we describe a simple numerical method for distinguishing cavity and channel loops from exterior loops. Cavity and channel loops are the ones inside the exterior loops (see Fig. 2 *b*). For the former group, when the individual arcs (more rigorously, the circles they are on) are traced counterclockwise, the whole loop is traced clockwise. For the latter group, when the individual arcs are traced counterclockwise, the whole loop is also traced counterclockwise. To classify a loop we imagine it is drawn on a piece of paper and we are looking down at the paper. The line integral over a loop

$$a = \oint \mathbf{n} \cdot (\mathbf{r} \times d\mathbf{l}) \tag{A1}$$

is positive if the loop is traced counterclockwise and negative if it is traced clockwise. In this expression $\mathbf{r}$ is a vector from an arbitrary but fixed point on the paper to the line segment $d\mathbf{l}$, and $\mathbf{n}$ is a unit vector perpendicular to and pointing out of rather than into the paper. The magnitude of a is twice the area of the loop.

As each loop is composed of pieces of arcs, we can express Eq. A1 as the sum of line integrals over the individual arcs. We then trace each arc counterclockwise. Suppose that the kth arc has starting and ending polar angles ϕ_{sk} and ϕ_{ek} on the circle centered at (x_{ck}, y_{ck}) with radius r_{ck}. Let the starting point of the vector $\mathbf{r}$ be at the origin; then when the ending point of $\mathbf{r}$ has a polar angle on the circle and the line segment $d\mathbf{l}$ spans $d\phi$ in the polar angle, we have

$$\mathbf{n} = (0, 0, 1), \tag{A2a}$$

$$\mathbf{r} = (x_{ck} + r_{ck} \cos \phi, y_{ck} + r_{ck} \sin \phi, 0), \tag{A2b}$$

and

$$d\mathbf{l} = (-r_{ck} \sin \phi \, d\phi, r_{ck} \cos \phi \, d\phi, 0). \tag{A2c}$$

The line integral over the kth arc now can be evaluated:

$$\begin{aligned} a_k = {} & r_{ck}^2(\phi_{ek} - \phi_{sk}) + x_{ck} r_{ck}(\sin \phi_{ek} - \sin \phi_{sk}) \\ & - y_{ck} r_{ck}(\cos \phi_{ek} - \cos \phi_{sk}). \end{aligned} \tag{A3}$$

The contributions from all the arcs of the loop are then summed. If the result is positive then the loop is an exterior one, otherwise it belongs to a cavity or a channel.

I thank Attila Szabo and Robert Zwanzig for many stimulating discussions. I am also grateful to Michael Gilson and one of the reviewers for informing me of the work of Juffer et al. and of Yoon and Lenhoff.

REFERENCES

Abramowitz, M., and I. A. Stegun. 1964. Handbook of Mathematical Functions. U.S. Department of Commerce, Washington, DC.

Davis, M. E., and J. A. McCammon. 1990. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* 90:509–521.

Jackson, J. D. 1962. Classical Electrodynamics. John Wiley and Sons, Inc., New York.

Juffer, A. H., E. F. F. Botta, B. A. M. van Keulen, A. van der Ploeg, and H. J. C. Berendsen. 1991. The electric potential of a macromolecule in a solvent: a fundamental approach. *J. Comp. Phys.* 97:144–171.

Klapper, I., R. Hagstrom, R. Fine, K. Sharp, and B. Honig. 1986. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modification. *Proteins Struct. Funct. Genet.* 1:47–59.

Louie, G. V., and G. D. Brayer. 1990. High-resolution refinement of yeast iso-1-cytochrome *c* and comparisons with other eukaryotic cytochromes *c*. *J. Mol. Biol.* 214:527–555.

McCammon, J. A., P. G. Wolynes, and M. Karplus. 1979. Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry.* 18:927–942.

Northrup, S. H., M. R. Pear, J. D. Morgan, J. A. McCammon. 1981. Molecular dynamics of ferrocytochrome *c*. Magnitude and anisotropy of atomic displacements. *J. Mol. Biol.* 153:1087–1109.

Northrup, S. H., J. O. Boles, and J. C. L. Reynolds. 1987. Electrostatic effects in the Brownian dynamics of association and orientation of heme proteins. *J. Phys. Chem.* 91:5991–5998.

Pelletier, H., and J. Kraut. 1992. Crystal structure of a complex between electron transfer partners, cytochrome *c* peroxidase and cytochrome *c*. *Science (Washington, DC).* 258:1748–1755.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1986. Numerical Recipes. Cambridge University Press, Cambridge.

Wang, J., J. M. Mauro, S. L. Edwards, S. J. Oatley, L. A. Fishel, V. A. Ashford, N-h. Xuong, and J. Kraut. 1990. X-ray structures of recombinant

yeast cytochrome *c* peroxidase and three heme-cleft mutants prepared by site-directed mutagenesis. *Biochemistry.* 29:7160–7173.

Warwicker, J., and H. C. Watson. 1982. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.* 157:671–679.

Yoon, B. J., and A. M. Lenhoff. 1990. A boundary element method for molecular electrostatics with electrolyte effects. *J. Comp. Chem.* 11: 1080–1086.

Yoon, B. J., and A. M. Lenhoff. 1992. Computation of the electrostatic interaction energy between a protein and a charged surface. *J. Phys. Chem.* 96:3130–3134.

Zauhar, R. J., and R. S. Morgan. 1985. A new method for computing the macromolecular electric potential. *J. Mol. Biol.* 186:815–820.

Zauhar, R. J., and R. S. Morgan. 1988. The rigorous computation of the molecular electric potential. *J. Comp. Chem.* 9:171–187.

Zauhar, R. J., and R. S. Morgan. 1990. Computing the electric potential of biomolecules: application of a new method of molecular surface triangulation. *J. Comp. Chem.* 11:603–622.

Zhou, H-X. 1993. Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys. J.* 64:1711–1726.