New Computational Models for Electrostatics of Macromolecules in Solvents

Jianhua Dai[1], Igor Tsukerman[1], Alexander Rubinstein[2], and Simon Sherman[2]

[1]Department of Electrical and Computer Engineering, The University of Akron, Akron, OH 44325-3904 USA
[2]Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198-6805 USA

Electrostatic fields of macromolecules (e.g., protein molecules) in solvents are often described by the Poisson–Boltzmann equation. This paper introduces two substantial amendments to the electrostatic model: first, the effective dielectric permittivity of the aqueous solvent layer on the molecular surface is drastically different from its bulk value of ∼80 and, second, the recently developed flexible local approximation methods produce different schemes with much higher accuracy than the classical ones.

***Index Terms*—Electrostatics, finite-difference (FD) schemes, flexible local approximation, macromolecules, Poisson–Boltzmann equation (PBE), proteins.**

I. Introduction

ELECTROSTATIC interactions are critical for protein structure and function [1]. Several computational techniques have been employed for such interactions: the finite-difference (FD) method [2], the finite-element method (FEM) [3], and the boundary-element method (BEM) [4], [5]. The recently developed flexible local approximation method (FLAME) [6], [7] incorporates accurate local approximations of the solution into the difference scheme and often yields higher accuracy on simple Cartesian grids than classical FD schemes and even than FEM with its complex meshes.

The electrostatic potential u in the solvent, under the assumptions of the mean-field theory, is governed by the Poisson–Boltzmann equation (PBE)

$$\nabla \cdot \varepsilon_s \nabla u = -\sum_{\alpha} n_\alpha q_\alpha \exp\left(-\frac{q_\alpha u}{k_B T}\right) \tag{1}$$

where summation is over all species of ions present in the solvent, n_α is volume concentration of species α in the bulk, q_α is the charge of species α, ε_s is the (absolute) dielectric permittivity of the solvent, k_B is the Boltzmann constant, and T is the absolute temperature. The right-hand side of the PBE reflects the Boltzmann distribution of microions. Inside the molecule, the potential satisfies the Poisson equation with atomic charges acting as point-like sources, and the standard boundary conditions apply on the molecule-solvent boundary. The effective permittivity of protein molecules is much smaller than that of the solvent.

When the term $q_\alpha u/(k_B T) \ll 1$, the PBE can be linearized by retaining the first two terms of the Taylor series of the exponential part [8]. Due to the electroneutrality, $\sum_\alpha n_\alpha q_\alpha = 0$, the linearized PBE becomes

$$\nabla^2 u - \kappa^2 u = 0 \quad \kappa^2 = \sum_{\alpha} \frac{n_\alpha q_\alpha^2}{k_B T \varepsilon_s}. \tag{2}$$

The computational model in this paper has two significantly new features. First, on the theoretical side, the effective dielectric constant in a thin layer near the molecule is about an order of magnitude lower than the bulk value of $\varepsilon_s \approx 80\varepsilon_0$ [9]. This significantly reduces the screening field effect and strongly increases the electrostatic interaction energy at the molecular interface. Second, on the numerical side, the recently developed FLAME method [6], [7] is applied. FLAME is a new FD calculus that incorporates accurate local approximations of the solution into the difference scheme. As the numerical test example (Section V) demonstrates, FLAME schemes yield much higher accuracy than the classical ones.

The FEM may be an alternative, but the complicated geometry of molecular surfaces makes finite-element mesh generation and solution very difficult [3]. FD schemes are, therefore, much more widely used in practice for this type of problem [2], and FLAME may offer a significant accuracy improvement in FD simulations.

II. Model and Formulation

The model of macroscopic continuum electrostatics is one of the most widely used to evaluate electrostatic interactions in the protein-solvent systems (e.g., [2], [10], [11], and [14]). In the framework of this model, in the classical approximation, the protein medium the protein medium is considered as a low-dielectric region (the dielectric constant of ∼2–4 [1]) immersed in a solvent with a high dielectric constant (∼80). Recently, it was shown in [9] that, by considering the dynamic solvent microstructure (particularly the correlated orientational Debye polarization of the water molecules) in the continuum model, sophisticated electrostatic effects at the protein-solvent interface can be revealed. These effects are impossible to represent in the classical framework (solvent as a homogeneous dielectric continuum).

In [9], the concept of nonlocal electrostatics and phenomenological theory of the polar solvent was used. The electrostatic problem with the standard boundary conditions was solved for a planar interface in terms of the Fourier transforms $\epsilon_1(k)$ and

Digital Object Identifier 10.1109/TMAG.2006.890959

$\epsilon_1(k)$ of the dielectric functions with spatial dispersion characterizing the bulk properties of the two condensed media in contact. The protein-like medium was considered as a uniform dielectric with a low dielectric constant $\epsilon_1(k) = \epsilon_1 = 4$. The orientational Debye polarization in aqueous solvent (determined by the hindered rotations of water dipoles due to hydrogen-bonding chains in bulk water) was modeled by the simple approximation of the dielectric function $\epsilon_2(k)$ for water

$$\epsilon_2(k) = \epsilon_* + \frac{\epsilon_s - \epsilon_*}{1 + (Lk)^2 \epsilon_s/\epsilon_*} \tag{3}$$

where $\epsilon_* = 6$ and $\epsilon_s = 78.3$ are short- and long-wavelength dielectric constants of the solvent at room temperature; L is the correlation length of water dipoles, proportional to the characteristic length of the hydrogen-bonding network of water molecules (∼5 Å).

In [9], for short-range pair-wise electrostatic interactions (inter-charge distances $< L$) in protein in close proximity ($< L$) to the interface, the value of the effective dielectric function was found to be very low, $\sim (\epsilon_1 + \epsilon_*)/2$, compared to $\sim (\epsilon_1 + \epsilon_s)/2$, determined by the classical model of the interface ($\epsilon_1 = 4$ and $\epsilon_* = \epsilon_s \approx 80$).

From the physical standpoint, the previous effect suggests the presence of an interfacial solvent layer with a low effective dielectric constant. The thickness of this layer is comparable with the correlation length L. The effective dielectric permittivity of the layer in close proximity to the dielectric boundary is determined by the short-wavelength dielectric constant of the bulk phase of the solvent ($\epsilon_* = 6$), which is much smaller than the bulk value, ϵ_s. This model is consistent with the experimental data on partially structured boundary water layer ("dynamically ordered water") on the surface of the protein native structure [13]. Overall, the data obtained in [9] suggests that the low-dielectric solvent layer on the protein surface is a critical factor that determines the electrostatic fields in the vicinity of the protein-solvent interface. The significant decrease in the screening of the field at the protein-solvent interface has been underestimated in traditional electrostatic models.

In this paper, we extend the model of [9] to 3-D systems. The full model of the solvent includes nonlocal interactions and is, therefore, computationally demanding. Here, we use a practical simplification of this model, with three dielectric layers (Fig. 1). The first layer is the protein itself. The second one is a layer of the solvent (∼6–8 Å) adjacent to the protein molecule; the dielectric constant of this layer is determined by the short-wavelength dielectric constant of the solvent $\varepsilon_* = 6$ [9]. Finally, the third layer represents the bulk phase of the solvent, with the relative dielectric constant of ∼80.

This paper is an initial step for testing and validation of new finite-difference methods for the heterogeneous electrostatics of macromolecules in solvents. Following similar analysis of conventional FD schemes by Gilson *et al.* [11], the protein molecule is simplified to a "particle" with radius 30 Å. The thickness of the intermediate dielectric layer in the solvent was set to 6 Å. Potential u in these two layers is governed by the electrostatic equation

$$\nabla \cdot \varepsilon \nabla u = \sum_{\alpha} q_\alpha \delta(\mathbf{r} - \mathbf{r}_\alpha) \tag{4}$$

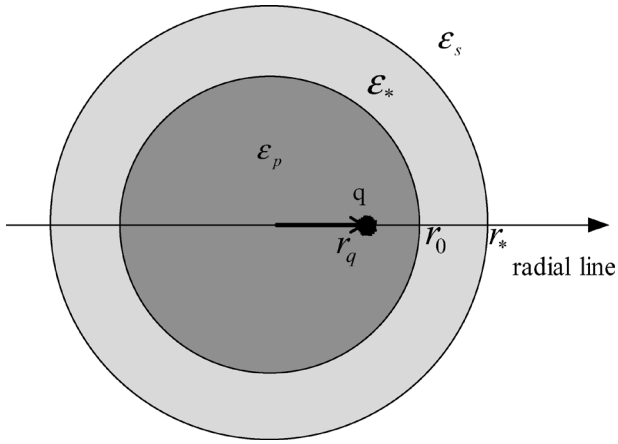

Fig. 1. 2-D representation of the model problem with three layers.

where q_α are the point charges corresponding to the protein atoms at locations $\mathbf{r}_\alpha$ inside the protein. The right-hand side of (4) in the intermediate layer of the solvent is zero. In the outer solvent layer, the potential is assumed to satisfy the PBE; in this paper, this equation is linearized (2), with the Debye–Hückel length of 8 Å corresponding to the physiological ionic strength.

The standard boundary conditions for the continuity of the electrostatic potential and for the normal component of the electric flux density apply at all dielectric interfaces. (If the derivatives in the electrostatic equations are understood in the sense of distributions, the boundary conditions are implied by the equations themselves.)

III. FLAME

FLAME is a recently developed generalized FD method [6], [7]. It has proven successful in a variety of applications, e.g., electrostatic multiparticle problems, the PBE, electromagnetic wave propagation, and scattering. The method incorporates local analytical approximations of the solution into the difference schemes in a simple way [6], [7].

For the protein problem, a regular Cartesian grid is introduced in the computational domain Ω. To construct a FLAME scheme over a given grid stencil, one chooses a set of basis functions satisfying equations (2) and (4), with proper boundary conditions at the dielectric interfaces. These basis functions can be constructed using spherical harmonics

$$\psi_{mn} = P_n^m(\cos\theta)\exp(\mathrm{i}m\phi) r^n \quad \text{inside layer 1} \tag{5}$$

$$\psi_{mn} = P_n^m(\cos\theta)\exp(\mathrm{i}m\phi)(c_{mn} r^n + d_{mn} r^{-n-1}) \quad \text{inside layer 2} \tag{6}$$

and

$$\psi_{mn} = P_n^m(\cos\theta)\exp(\mathrm{i}m\phi)(f_{mn} j_n(\mathrm{i}kr) + g_{mn} n_n(\mathrm{i}kr) \tag{7}$$

inside layer 3. Here, $j_n(z) = (\pi/(2z))^{1/2} J_{n+1/2}(z)$ and $n_n(z) = (\pi/(2z))^{1/2} Y_{n+1/2}(z)$ are spherical Bessel functions of the first and second kind, respectively. They are expressible in terms of the hyperbolic sine and cosine functions and are, therefore, fairly easy to work with. The coefficients c_{mn}, d_{mn}, f_{mn}, and g_{mn} are found from the interface boundary conditions in a standard way [6].

The numerical solution over the ith grid stencil in FLAME is a linear combination of the local basis functions

$$u_h^{(i)} = \sum_{\alpha=1}^{m^{(i)}} c_\alpha^{(i)} \psi^{(i)} \tag{8}$$

where $c_\alpha^{(i)}$ are some coefficients. The FLAME scheme is [7]

$$\underline{s}^{(i)} \in \mathrm{Null}\left(N^{(i)T}\right) \tag{9}$$

where matrix $N^{(i)}$ contains the nodal values of the basis functions: $N_{\beta\alpha}^{(i)} = \psi_\alpha(r_\beta)$, $\alpha = 1, 2, \ldots, m^{(i)}$, $\beta = 1, 2, \ldots, M^{(i)}$, and r_β are the position vectors of the stencil nodes. In this paper, we consider 7-point and 19-point FLAME schemes; their respective basis sets contain 6 and 18 spherical harmonics, which leads to the null space of dimension one in (9).

For the inhomogeneous equation of the general form

$$Lu = f \tag{10}$$

the FLAME scheme is constructed by splitting the solution as

$$u^{(i)} = u_0^{(i)} + u_f^{(i)} \tag{11}$$

where $u_0^{(i)}$ is the solution of the homogeneous equation and $u_f^{(i)}$ is a particular solution of the inhomogeneous equation. With this splitting, the inhomogeneous FLAME scheme is [6]

$$s^{(i)T} u^{(i)} = s^{(i)T} u_f^{(i)}. \tag{12}$$

Not surprisingly, the coefficients of the scheme, and hence, the system matrices, for the homogeneous and inhomogeneous equations are the same. The right-hand side of the inhomogeneous problem is obtained by applying the FLAME scheme to the particular solution $u_f^{(i)}$.

The FLAME solution of the inhomogeneous equation is

$$u^{(i)} = \sum_{\alpha=1}^{m^{(i)}} c_\alpha^{(i)} \psi^{(i)} + u_f^{(i)} \tag{13}$$

which in practice is used as an interpolation formula for the potential and for the electric field (after differentiation of the basis functions).

IV. Treatment of Point Charges

In traditional FD methods, point charges are usually "assigned" to a set of nearby grid nodes; nodes that are closer to the charge get a greater portion of that charge assigned to them. This charge discretization introduces additional errors, especially when the charge is located close to the protein boundary [12]. In contrast, for the potential splitting previously described, there is *no* numerical error due to the potential of point charges, as this potential is represented analytically and exactly. The numerical results in Section V prove this.

The particular solution $u_f^{(i)}$ in the vicinity $\Omega^{(i)}$ of a point charge is constructed as follows. If $\Omega^{(i)}$ lies entirely inside the protein molecule, $u_f^{(i)}$ can simply be calculated by Coulomb's law. (Note that this particular solution is purely local, defined only in $\Omega^{(i)}$, and need not satisfy any boundary conditions outside this local region.) If $\Omega^{(i)}$ contains an interface boundary between two dielectric regions, the particular solution is computed by matching the spherical harmonic expansions in these regions. (The harmonic expansion of the potential of a point charge is well known in electromagnetic theory; see, e.g., [15].)

V. Numerical Results

In the test examples that follow, a single point charge inside the molecule is the source of the field. The location of the charge may vary. The relative error in the numerical calculations is defined as

$$\mathrm{err}(x,y,z) = \frac{|u_{\mathrm{num}}(x,y,z) - u_{\mathrm{an}}(x,y,z)|}{|u_{\mathrm{num}}(x,y,z)|} \times 100\% \tag{14}$$

where u_{num} is the numerical potential and u_{an} is the analytic result via a series of spherical harmonics, truncated in our analysis at the terms with the magnitude of less than 10^{-5}. To eliminate the numerical error associated with the approximation of boundary conditions, the exact (theoretical) Dirichlet condition is applied on the domain boundary.

First, let the point charge be located inside the protein, 6 Å from the surface, at $(x, y, z) = (0, 0, 24)$. (All distances are measured in angstroms.) The mesh size is 1.5 Å in each of the three directions.

The relative error as a function of distance from the center of the protein molecule is evaluated as the root mean square (rms) over 100 evenly distributed sampling points on a spherical surface of a given radius.

Fig. 2 shows that the accuracy of the FLAME schemes is much higher than that of standard FD with charge assignment. Note that the accuracy of FLAME is high even in the area around the point charge, where the potential is singular; this is because the potential of the charge is represented in FLAME exactly, via the potential splitting described before. The average rms error of the standard FD is about 8.96% compared to 1.98% for the 7-point FLAME scheme and 0.32% for the 19-point FLAME scheme. The 19-point scheme yields higher accuracy, at the expense of the higher density of the system matrix.

In the second example, the point charge is very close—at 1 Å to the protein surface, at $(x, y, z) = (0, 0, 29)$. The proximity of the charge to the surface makes the problem much more difficult computationally. For high accuracy in FLAME, a fairly large number of spherical harmonics must be retained in the expansion of the particular solution $u_f^{(i)}$.

Other geometric and physical parameters (the mesh size, permittivities, etc.) remain the same as in the previous case. The rms errors as a function of the distance from the origin are plotted in Fig. 3. FLAME again provides much higher accuracy than the standard FD. The average rms error is 1.70% for the 7-point FLAME scheme and 0.54% for the 19-point FLAME scheme, as compared with 12.62% for standard FD. The accuracy of the 7-point FLAME scheme deteriorates slightly in the solvent, while the 19-point scheme is uniformly very accurate throughout the domain.

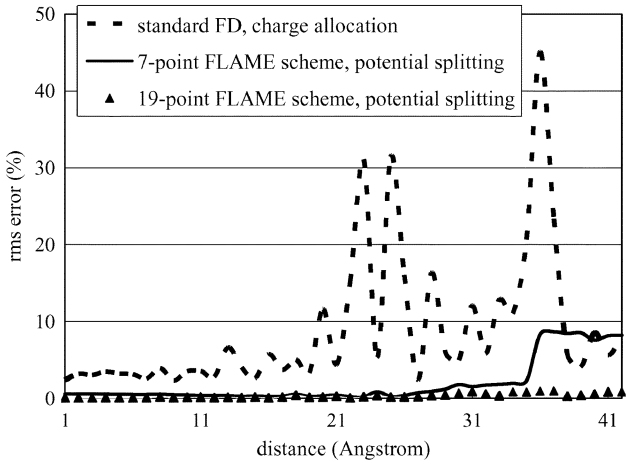

Fig. 2. Errors as a function of distance from the center of the molecule. The radius of the molecule is 30 Å and the charge is located 6 Å from its surface.

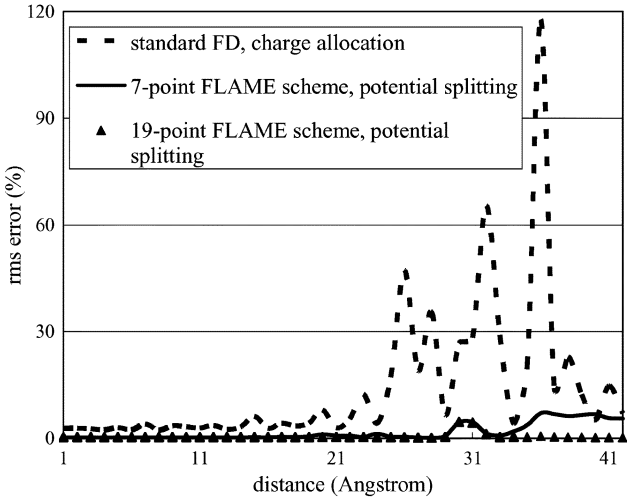

Fig. 3. Errors as a function of the distance from the origin. The protein has radius 30 Å and the charge is located 1 Å from its surface. 50 spherical harmonics are retained in the expansion of $u_f^{(i)}$.

VI. Conclusion

This paper introduces a new computational model for the electrostatics of protein molecules in solvents. Three dielectric layers are present: interior of the molecule with a low dielectric constant of ∼2–4; a thin layer of solvent near the molecule with the dielectric constant of ∼6; and the bulk of the solvent with the dielectric constant of ∼80. The electrolyte in the third layer is described by the linearized PBE.

Applications of the FEM to electrostatics of macromolecules [3] are encumbered by the enormous complexity of the meshes. An alternative approach, the FD method, is more practical for this type of problem [2]. In this paper, the new FD calculus of FLAME is applied. For point charges as sources of the field, potential splitting is used instead of the conventional charge allocation to the grid. The numerical results confirm much higher accuracy of FLAME schemes, as compared with the standard FD analysis.

In the presence of point sources, the accuracy of FLAME depends on the particular solution in the potential splitting. For point charges located sufficiently far away from the molecular surface, Coulomb's law may be used as the (local) particular solution. For charges close to the surface, multiple spherical harmonics must be retained in the particular solution. This improves the accuracy dramatically, as compared to the standard finite difference method with charge allocation to grid nodes.

Acknowledgment

The work of J. Dai and I. Tsukerman was supported in part by the National Science Foundation under Research Awards 0304453 and 9812895. The Bioinformatics Core Facility of the Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center used in this work was partially supported by the Cancer Center Support under Grant P30 CA36727.

References

[1] T. Simonson, "Electrostatics and dynamics of proteins," *Rep. Prog. Phys.*, vol. 66, no. 5, pp. 737–787, 2003.

[2] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig, "Rapid grid-based construction of the molecular-surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects," *J. Comput. Chem.*, vol. 23, pp. 128–137, 2002.

[3] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: Application to microtubules and the ribosome," *PNAS* vol. 98, no. 18, pp. 10037–10041, 2001 [Online]. Available: http://www.pnas.org/cgi/content/abstract/98/18/10037

[4] B. J. Yoon and A. M. Lenhoff, "A boundary element method for molecular electrostatics with electrolyte effects," *J. Comp. Chem.*, vol. 11, no. 9, pp. 1080–1086, 1990.

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

[6] I. Tsukerman, "Electromagnetic application of a new finite-difference calculus," *IEEE Trans. Magn.*, vol. 41, no. 7, pp. 2206–2225, Jul. 2005.

[7] ——, "A class of difference schemes with flexible local approximation," *J. Comput. Phys.*, vol. 211, no. 2, pp. 659–699, 2006.

[8] F. Fogolari, P. Zuccato, G. Esposito, and P. Viglino, "Biomolecular electrostatics with the linearized Poisson–Boltzmann equation," *J. Biophys.*, vol. 76, pp. 1–16, 1999.

[9] A. Rubinstein and S. Sherman, "Influence of the solvent structure on the electrostatic interactions in proteins," *Biophys. J.*, vol. 87, pp. 1544–1557, 2004.

[10] M. Gilson, A. Rashin, R. Fine, and B. Honig, "On the calculation of electrostatic interactions in proteins," *J. Mol. Biol.*, vol. 183, pp. 503–516, 1985.

[11] M. Gilson, K. Sharp, and B. Honig, "Calculating the electrostatic potential of molecules in solution: Method and error assessment," *J. Comp. Chem.*, vol. 9, pp. 327–335, 1988.

[12] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig, "Focusing of electric fields in the active site of Cu–Zn superoxide dismutase," *Proteins*, vol. 1, pp. 47–59, 1986.

[13] S. K. Pal and A. H. Zewail, "Dynamics of water in biological recognition," *Chem. Rev.*, vol. 104, pp. 2099–2123, 2004.

[14] A. Papazyan and A. Warshel, "Continuum and dipole-lattice models of solvation," *J. Phys. Chem. B*, vol. 101, pp. 11254–11264, 1997.

[15] W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism*. Reading, MA: Addison-Wesley, 1962.

Manuscript received May 1, 2006 (e-mail: jd35@uakron.edu).