Daniel A. Beard
Tamar Schlick
Department of Chemistry and Courant Institute of Mathematical Sciences, New York University and Howard Hughes Medical Institute, 251 Mercer Street, New York, NY 10012

Received 2 May 2000; accepted 28 June 2000

Modeling Salt-Mediated Electrostatics of Macromolecules: The Discrete Surface Charge Optimization Algorithm and Its Application to the Nucleosome

Abstract: *Much progress has been achieved on quantitative assessment of electrostatic interactions on the all-atom level by molecular mechanics and dynamics, as well as on the macroscopic level by models of continuum solvation. Bridging of the two representations—an area of active research—is necessary for studying integrated functions of large systems of biological importance. Following perspectives of both discrete (*N*-body) interaction and continuum solvation, we present a new algorithm, DiSCO (Discrete Surface Charge Optimization), for economically describing the electrostatic field predicted by Poisson–Boltzmann theory using a discrete set of Debye–Hückel charges distributed on a virtual surface enclosing the macromolecule. The procedure in DiSCO relies on the linear behavior of the Poisson–Boltzmann equation in the far zone; thus contributions from a number of molecules may be superimposed, and the electrostatic potential, or equivalently the electrostatic field, may be quickly and efficiently approximated by the summation of contributions from the set of charges. The desired accuracy of this approximation is achieved by minimizing the difference between the Poisson–Boltzmann electrostatic field and that produced by the linearized Debye–Hückel approximation using our truncated Newton optimization package. DiSCO is applied here to describe the salt-dependent electrostatic environment of the nucleosome core particle in terms of several hundred surface charges. This representation forms the basis for modeling—by dynamic simulations (or Monte Carlo)—the folding of chromatin. DiSCO can be applied more generally to many macromolecular systems whose size and complexity warrant a model resolution between the all-atom and macroscopic levels.* © 2000 John Wiley & Sons, Inc. Biopoly 58: 106–115, 2001

Keywords: *electrostatics; Poisson–Boltzmann theory; nucleosome core particle; chromatin*

INTRODUCTION

Salt-mediated electrostatic interactions play a key role in governing the structural and dynamic details of many important macromolecular processes, such as the mechanism of ion channels and the folding of nucleoprotein complex chromatin. Effective use of the emerging wealth of data on sequences and struc-

Correspondence to: Tamar Schlick; schlick@nyu.edu
Contract grant sponsor: National Science Foundation (NSF) and National Institutes of Health (NIH)
Contract grant number: ASC-9157582 (NSF) and R01 GM55164
Biopolymers, Vol. 58, 106–115 (2001)
© 2000 John Wiley & Sons, Inc.

tures of biomolecules requires strategies for distilling detailed information into a minimal representation that captures the important features of a given system. For example, the cellular-level organization of DNA, on the scale of thousands of base pairs and more, is modeled using a uniformly charged elastic polymer representation for DNA, with the DNA immersed in an electrolytic viscous solvent. It is far more difficult to model the proteins to which DNA is bound in living systems, and that regulate not only the macroscopic structure of the DNA but also major biochemical pathways associated with the genetic material. The sizes of such systems demand biophysical descriptions in the spirit of polymer-level models of DNA. Yet these systems are much less regular in terms of shape and charge distribution than simple DNA and require a more sophisticated treatment. In this paper we introduce and apply to the nucleosome core particle (see Figure 1 and Ref. 1) our method DiSCO (Discrete Surface Charge Optimization) for producing an optimized discrete *N*-body Debye–Hückel potential to match the electric field predicted by the nonlinear Poisson–Boltzmann equation. This economical description of the electrostatics associated with a solvated macromolecule can be incorporated into a macrolevel model, such as a model for the nucleoprotein complex chromatin, the system that motivated the present study. More generally, DiSCO allows integrating atomic-level details of a biomolecule into an accurate biophysical description of a system too large to treat on the atomic scale.

In our application of DiSCO to the crystallographically solved nucleosome core particle,[1] we show that the DiSCO approximation is accurate for distances within one Debye length away from the surface of macromolecule (e.g., 14 Å at a salt concentration of 0.05*M*). Specifically, a 277-charge Debye–Hückel approximation is accurate to within 10% error at distances greater that 10 Å from the surface of the core particle. As the number of effective charges and the distance from the surface of the macromolecule is increased, the description becomes more accurate. This accuracy at a range of salt concentrations warrants incorporation of this charged nucleosome model into a macrolevel model for the structure and dynamics of the chromatin fiber,[2] an application limited to date by an accurate treatment of the core particle charge. More generally, our approach is applicable to many other biomolecular systems whose size and complexity warrant a model resolution between the all-atom and macroscopic levels.

In the next section we review several approaches to modeling electrostatic interactions in biomolecular systems, including the application of Poisson–Boltzmann theory to solvated macromolecules. In the Methods section we outline the theory and methods behind the *N*-body parameterization. Specifically, a rudimentary outline of Poisson–Boltzmann theory applied to macromolecules is presented. The details involved in constructing the *N*-body Debye–Hückel parameterization are presented next. In the Results section, we detail the application of the DiSCO model to describe the electrostatics of the nucleosome core particle. Our work combined with that of Stigter for the representation of salt-mediated electrostatic effects of biopolymers now makes possible a new line of applications to chromatin folding.[2]

BACKGROUND

Modeling Challenges on Many Levels

Significant advances over the past decade have made the calculation of long-range electrostatic interactions in molecular dynamics (MD) simulations more efficient. Namely, fast multipole and particle mesh Ewald techniques (see Darden et al., 1999,[36] and refs. therein) have reduced the associated $\mathcal{O}(N^2)$ work for a system of *N* atoms to nearly linear complexity. Still, when *large systems* are modeled—such as a solvated protein embedded in a phospholipid bilayer (91,000 atoms; 45,000 flexible)[3]—or when long simulations are performed—such as a microsecond trajectory for a 36-residue peptide (12,000 atoms)[4]—the nonbonded interactions are simplified via cutoffs, even when dedicated supercomputing resources are available.

Another approach to simplifying the atom-level *N*-body calculation is to exclude explicit treatment of the solvent and to construct a pairwise interaction model that considers a dielectric cavity surrounding the molecular charges. Such generalized Born potentials, which can be modified to approximate salt-screening effects, have proven useful at describing thermodynamic properties of the solvation of macromolecules, i.e., consistent with those generated from detailed simulations based on explicit water models.[5,6] However, a realistic treatment of solvent hydrodynamics is missing from these studies.

This limitation can be addressed by invoking continuous approximations, not only for solvent electrostatics, but also for solvent hydrodynamics. For example, interactions of solvent molecules with the macromolecular system can be modeled as a combination of stochastic and frictional forces. The correlation structure of the stochastic forces is related to the form of the frictional interaction model, which is included to mimic the continuum-level mechanics of the solvating fluid.[7] A popular application of this approach is in the simulation of polymer-level models

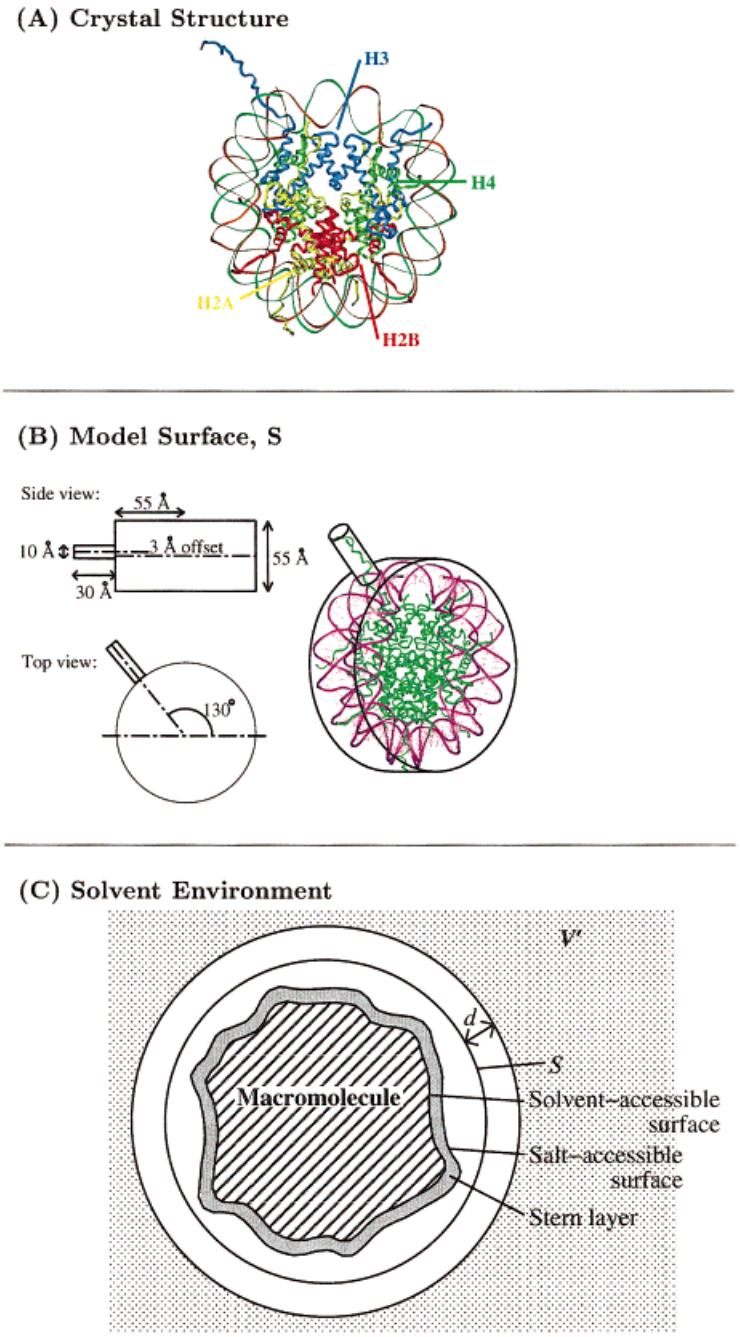

FIGURE 1 (A) Crystal structure of nucleosome core particle from Luger et al.[1] A left-handed supercoil of 146 bp of DNA is wrapped around a protein core composed of an octamer of histone proteins (H2A, H2B, H3, and H4). (B) Sketch of the surface S used to enclose the core particle. The surface is composed of a disk and a cylinder: A disk of 55 Å radius and height of 55 Å encloses the main body of the core particle; a cylinder of 5 Å and 30 Å height encloses the tail of the H3 histone. (C) The relationship between the surface S and the region V' on the exterior of the macromolecule. The Stern layer lies between the solvent-accessible and the salt-accessible surfaces of the macromolecule. The surface S, upon which the N Debye–Hückel charges are arranged, encloses the macromolecule. At a distance d beyond S lies the region V' in which the N-body approximation is compared to the field predicted by the PBE.

of DNA,[8–11] using Brownian dynamics (BD).[12] These formulations employ an *N*-body representation of the hydrodynamic interactions among the subunits of the DNA model, specified by a $3N \times 3N$ matrix, such as the Rotne–Prager tensor,[13] which relates the velocity of a given subunit to the hydrodynamic forces transmitted to all the particles in the system.

The Poisson–Boltzmann (PB) Approach

For macroscopic DNA models, electrostatic interactions are typically described by an *N*-body potential that arises from the mean field PB treatment of dissolved ions and solvent. In PB theory,[14–16] the electrostatic field is governed by Gauss' Law for electrostatics.[17] Solvating water is modeled as an inhomogeneously distributed continuous dielectric medium, and dissolved ions are treated as a continuous concentration field, distributed according to the principles of statistical mechanics (see Methods). These assumptions are expressed mathematically in the well-known Poisson-Boltzmann equation (PBE), which is nonlinear and, in general, only solvable using approximate numerical methods.

PB theory can be used to treat the salt-screened electrostatic interactions between biomolecules and to calculate long-range forces used in simulating diffusion-controlled interactions,[18] such as in enzyme–substrate interactions. However, in such applications, the PBE must be reformulated and solved for each new geometry as the relative orientation of the molecules changes during the simulation. The large amount of computational work involved in this repeated PBE formulation can be reduced by superposing solutions to the linearized PBE (LPBE) for molecules that are sufficiently far apart. Since exact solutions for the LPBE in homogeneous media are easily obtained from the superposition of fundamental solutions for point charges, an approximation of the far-field solution to the PBE, using the field produced by a number of point sources for the LPBE, can be constructed. Since the biomolecular system is neither homogeneous nor linear, the values of the point sources, which are effective Debye–Hückel charges, must be varied to optimally reproduce the far zone electrostatic field.

N-Body Approaches Based on PB Theory

In the late 1970s, Stigter[19] pioneered such a Debye–Hückel approximation for DNA electrostatics. His approximation is based on solving the PBE for an infinitely long and uniformly charged DNA cylinder and matching the resulting solution in the far zone to a solution to the LPBE associated with a line of uniform charge density located at the center of the DNA cylinder. The nonlinear solution was possible in this case (without the aid of modern computers!) because the cylindrical problem is formulated in terms of only one independent variable, radial distance. Stigter's work has made possible numerous applications of BD and Langevin dynamics to DNA systems of several thousand base pairs.[8,11,20,21]

Proteins present a more difficult subject for this Debye–Hückel approach to salt-mediated electrostatics because their geometry and charge distribution is much less regular than for DNA. Still, this representation can be generalized to irregularly shaped molecules by constructing the linear approximation from a number of discrete Debye–Hückel charges distributed on the surface of the biomolecule. The value assigned to each charge on the surface can be obtained by matching the Debye–Hückel electrostatic field to the field generated based on the nonlinear PBE.

For our general method DiSCO, the accuracy of the discrete Debye–Hückel approximation is assessed by formulating and minimizing an objective function that reflects the residual between the electrostatic field predicted by the PBE and the field associated with the Debye–Hückel charges. The minimum value of the residual depends on the number of point charges, the definition of the far zone (distance from the macromolecule where the linear approximation is valid), and the salt concentration.

A similar approach for finding an effective charge description for solvated macromolecules was introduced by Gabdoulline and Wade.[22] By representing the electrostatics of a protein with charges placed at the center of each residue head group and each charged side chain, they matched the resulting potential to that obtained from the linearized Poisson-Boltzmann equation. Our approach uses the electric field associated with the nonlinear Poisson–Boltzmann equation, rather than the potential from the linear equation, as the standard for evaluating the accuracy of the effective charge approximation. This makes it possible to treat the electrostatic forces associated with large highly charged biomolecules more directly. We also allow for a variable number of point charges to be distributed on the surface enclosing the macromolecule; this number can be adjusted based on desired accuracy and resolution, and balanced with the computational work entailed.

METHODS

Poisson–Boltzmann Theory

The PBE for an electrostatic potential ϕ as a function of position $\mathbf{r}$ for a solvated system containing monovalent ions is described as[14]

$$\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r})\sinh(\phi(\mathbf{r})) = -4\pi\rho(\mathbf{r})/k_BT \quad (1)$$

where $\epsilon(\mathbf{r})$ is the position-dependent permittivity (typically smaller in the interior macromolecular region) and $\rho(\mathbf{r})$ is the fixed charge density. The Debye–Hückel parameter $\bar{\kappa}(\mathbf{r})$ is defined as

$$\bar{\kappa}^2(\mathbf{r}) = \frac{8\pi e^2 N_A C_s}{k_BT} \quad (2)$$

in the solvent. Here N_A and k_B denote Avagadro's and Boltzmann's constants, respectively; e is the elementary charge, C_s is the bulk ion concentration in moles per liter, and T is the absolute temperature. In Equation (1), ϕ is a dimensionless quantity, $\phi = e\Phi/k_BT$, where Φ is the potential in units of electrostatic energy per unit charge.

The position dependency of $\bar{\kappa}^2(\mathbf{r})$ is determined by the volume of the system accessible by the electrolyte. In regions not available to salt, $\bar{\kappa}^2(\mathbf{r}) = 0$. Thus the Debye length κ^{-1}, or

$$\frac{1}{\kappa} = \frac{\epsilon^{1/2}(\mathbf{r})}{\bar{\kappa}(\mathbf{r})} \quad (3)$$

is effectively infinite on the interior of a macromolecule. The salt exclusion zone is usually assumed to extend beyond the interior of the macromolecule.

The PBE can be linearized by expanding the hyperbolic sine function as a Taylor series and retaining only the first-order term. The resulting LPBE approximation,

$$\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi(\mathbf{r}) - \bar{\kappa}^2\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r})/k_BT \quad (4)$$

is valid for relatively weak potentials $e\phi \ll 1$.

DiSCO—Construction of the *N*-Body Debye–Hückel Approximation

The N-body Debye–Hückel parameterization is constructed by optimizing the values of the effective charges according to the procedure detailed below:

1. *Solve the PBE.* Solve the PBE for a given three-dimensional macromolecular structure using atomic-level partial charges and the desired salt concentration. The resulting potential $\phi(\mathbf{r})$, defined for the numerical grid points $\{\mathbf{r}_i\}$, corresponds to the electrostatic field $\mathbf{E}(\mathbf{r}_i) = -\nabla\phi(\mathbf{r}_i)$. The DelPhi package developed by Honig et al.[16,23–26] or the UHBD package of McCammon and co-workers[18] can be used to obtain a solution to a finite difference approximation of the PBE on a Cartesian grid $\{\mathbf{r}_i\}$.
2. *Construct a Virtual Surface* S. The surface S encloses the macromolecule (see Figure 1) and is used to define the locations of the N Debye–Hückel charges.
3. *Calculate the Effective Charge on* S. Calculate the surface charge $\sigma(\mathbf{r})$ corresponding to the electrostatic field $\mathbf{E}(\mathbf{r})$ from

$$\sigma(\mathbf{r}_j) = \frac{\epsilon}{4\pi}\mathbf{n}_j \cdot \mathbf{E}(\mathbf{r}_j) \quad (5)$$

where $\sigma(\mathbf{r}_j)$ is the surface charge density, $\mathbf{r}_j$ is a point on the surface S, and $\mathbf{n}_j$ is the unit outward normal to S.

4. *Define the Debye–Hückel Charges.* The N Debye–Hückel charges $\{q_j^*\}$, located on S, are expressed in terms of the surface charge density $\sigma(\mathbf{r})$:

$$q_j^* = \sigma(\mathbf{r}_j)dS_j \quad (6)$$

where dS_j is the finite surface area associated with the point $\mathbf{r}_j$ on the surface S.

5. *Optimize the Residual Function.* The residual function

$$R(\tilde{\mathbf{E}}, \mathbf{E}) = \frac{1}{V'}\int_{V'} \frac{\|\mathbf{E} - \tilde{\mathbf{E}}\|}{\|\mathbf{E}\|}\, dV \quad (7)$$

or discretized version

$$R(\tilde{\mathbf{E}}, \mathbf{E}) = \frac{1}{N_v}\sum_{i \in V_i'} \frac{\|\mathbf{E}(\mathbf{r}_i) - \tilde{\mathbf{E}}(\mathbf{r}_i)\|}{\|\mathbf{E}(\mathbf{r}_i)\|} \quad (8)$$

is a measure of the difference between $\mathbf{E}$ (the field predicted by the PBE) and $\tilde{\mathbf{E}}$ (the field predicted by the Debye–Hückel equation). In these equations V' is a given volume on the exterior of the surface S, and V_i' is the set of grid points contained in V' (e.g., all grid points that lie within a certain range of distance from S; see Figure 1), and N_v is the number of grid points in V_i'. The field $\tilde{\mathbf{E}}(\mathbf{r}_i)$ is evaluated as the negative gradient of the Debye–Hückel potential generated by the N charges $\{q_j\}$

$$\tilde{\phi}(\mathbf{r}_i, \{q_j\}) = \sum_{j=1}^{N} \frac{q_j e^{-\kappa r_{ij}}}{\epsilon r_{ij}} \quad (9)$$

which is an exact solution to the LPBE, in the homogeneous solvent, away from the molecular boundary. Here ϵ is the permittivity of the solvent, and r_{ij} is the scalar distance between position vectors $\mathbf{r}_i$ and $\mathbf{r}_j$ (the locations of the ith and jth charges). The corresponding field is

$$\tilde{\mathbf{E}}(\mathbf{r}_i) = \frac{1}{\epsilon}\sum_{j=1}^{N} q_j e^{-\kappa r_{ij}}\left[\frac{\kappa}{r_{ij}^2} + \frac{1}{r_{ij}^3}\right](\mathbf{r}_i - \mathbf{r}_j) \quad (10)$$

The residual function is minimized by varying the charges charges $\{q_j\}$ (independent variables). For this minimization task, we use the efficient truncated Newton package developed by Schlick and co-workers.[27–29]

We have also considered an alternative approach for charge optimization using a residual R_ϕ based on the potential ϕ, rather than the electrostatic field $\mathbf{E}$ as above. We find that this alternative approach is less robust. Namely, when optimal charges based on R_ϕ are evaluated using the electric field residual of Eq. (8), the corresponding gradient is very large. This is not the case when optimized charges based on the electric field residual are used as an initial

guess for minimizing R_ϕ. This is likely because differences between the Debye–Hückel field and the field predicted by the PBE vary in space more smoothly for the electrostatic potential than for the electrostatic field.

RESULTS

The Nucleosome Core Particle and PBE Calculations

The beautiful nucleosome core crystal structure reported by Luger et al.[1] (PDB code 1A0I) is an octamer protein core (two copies each of H2A, H2B, H3, and H4; see Figure 1A) surrounded by 146 base pairs (bp) of left-handed supercoiled DNA. Figure 1B shows the virtual surface S used to enclose the nucleosome core. The surface is constructed from a disk and a slender cylinder. The disk (55 Å radius and 55 Å width) encloses the main body of the octamer and the wrapped DNA. The other cylinder (5 Å radius and 30 Å length) encloses the H3 tail, which extends between the gyres of the wrapped DNA.

The atomic partial charges are assigned using the AMBER force field developed by Cornell et al.[30] As discussed in Honig et al.,[16] the electrostatic permittivity ϵ is set to $2\epsilon_o$ on the interior and $80\epsilon_o$ on the solvent-accessible exterior of the biomolecule (ϵ_o is the free-space permittivity). The solvent-accessible region is determined as the set of all points on which a centered 1.4 Å radius sphere used to represent a water molecule does not overlap the van der Waals radii of the atoms making up the core particle. A similar procedure is adopted to determine the salt-accessible region. Our values of 2 Å radii for sodium and chloride ions result in an approximately 0.6 Å Stern layer, modeled as a high dielectric solvent ($\epsilon = 80\epsilon_o$) with no dissolved ions ($\bar{\kappa} = 0$).

Because the core particle is relatively highly charged (the total negative charge assigned is $-2234e$ while the total positive charge is 2036 e), a robust nonlinear solver is required for obtaining solutions to the PBE. The Qnifft DelPhi solver[31] is implemented with a grid of 65^3 points, the maximum allowable. Using the resolution 2.5 Å, the grid corresponds to a cubic domain of $160 \times 160 \times 160$ Å^3.

The "focusing" boundary condition for PBE calculations[24,32] is used to minimize edge effects in the numerical solution by performing calculations on successively smaller grids. For each successive iteration, the boundary of the computational domain lies within the solution obtained at the previous level of refinement. The boundary condition is set by interpolating the potential of the previous solution onto the boundary of the grid.

We performed calculations at various salt concentrations ranging between $C_s = 0.01M$ and $C_s = 0.10M$. The Grasp-generated[33] images in Figure 2 for $C_s = 0.01$ and $0.10M$ show that the DNA region adopts a mainly negative potential (red), while the surface of the histones includes regions of both negative and positive (blue) potential. The H3 tail adopts a mostly positive potential, due to the ubiquity of the basic amino acid lysine. At the higher salt, increased screening reduces the magnitude of the surface potential.

Generation of *N*-Body Potential by DiSCO

We distribute the surface point charges at six levels of resolution ($N = 62$, 79, 144, 199, 277, and 353) as shown in Figure 3. For each case, charges are arranged in a hexagonal pattern on each of the flat surfaces of the main disk with additional charges are arranged on the tail cylinder and the circular outer edge of the main disk. The figure indicates the number of charges distributed over each region—the top and bottom flat surfaces of the core disk, the circular outer edge of the disk, and the H3 tail body.

In optimizing the N-body potential, we introduce the parameter d to define V', the set of grid points over which the N-body approximation is assessed relative to the nonlinear PBE potential. The set V' is composed of all grid points for which the minimum distance between the grid point and the surface S is greater than or equal to d. The relationship between the macromolecular surfaces, the enclosing surface S, and the region V' is diagrammed in Figure 1B.

The set of initial charges $\{q_j^*\}$, the Debye–Hückel potential $\tilde{\phi}(\mathbf{r}_i, \{q_j^*\})$, and associated electrostatic field $\tilde{\mathbf{E}}(\mathbf{r}_i, \{q_j^*\})$, produces the residual $R(\tilde{\mathbf{E}}(\{q_j^*\}), \mathbf{E})$ for the initial guess (see Methods). For $C_s = 0.05M$, $N = 353$, and $d = 10$ Å, we find that the field generated by the initial guess set of Debye–Hückel charges $\{q_j^*\}$ differs from the field predicted by the nonlinear PBE by about 57% in the region V_i' ($R = 0.57$). The TNPACK minimizer reduces the gradient of the objective function by ten orders of magnitude in 50 min of CPU time on a Silicon Graphics R10000 processor, resulting in an optimized value of 6% [$R(\tilde{\mathbf{E}}(\{q_j\}) = 0.06$]. For $N = 62$ charges, the initial error of 60% is reduced to 21%, but requires less than 1 min of CPU time. Figure 4 shows the optimal value of R as a function of salt concentration for various values of N.

Interestingly, accuracy does not vary monotonically with salt concentration. For example, at $N = 199$ the minimum error R is obtained at $C_s = 0.02M$. This minimum exists because at lower salt

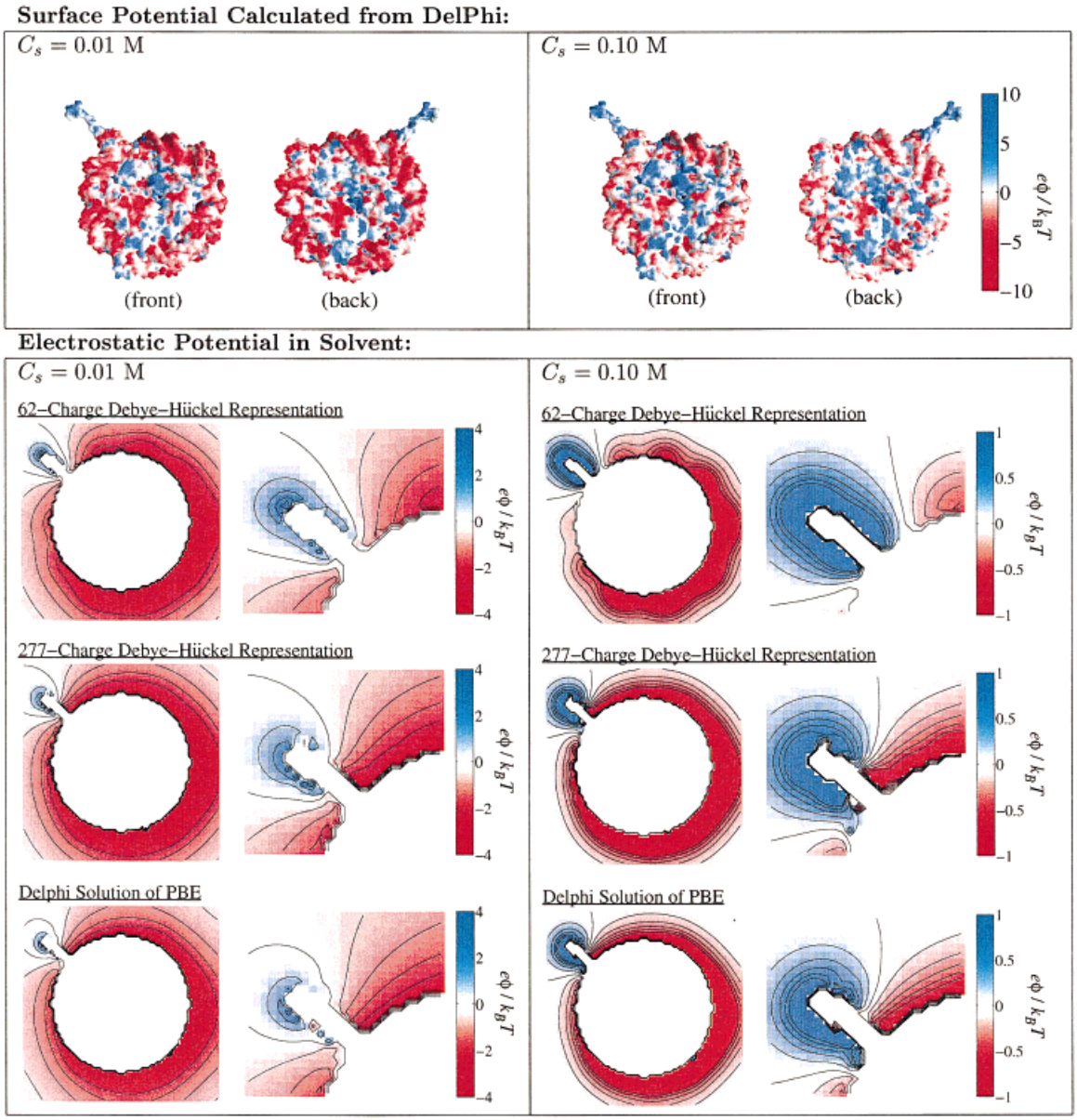

FIGURE 2 Core particle electrostatic potential in two salt environments. Upper panel: Electrostatic potential on the surface of the nucleosome core particle, as predicted by the finite-difference approximation of the PBE and rendered by Grasp[33] at $C_s = 0.01M$ (left) and $C_s = 0.10M$ (right). Lower panel: *N*-body vs PBE-computed electrostatic potentials in the solvent surrounding the core particle, as predicted by our model with $N = 62$ and 277 charges vs that predicted by the finite difference solution to the PBE, at $C_s = 0.01M$ (left) and $C_s = 0.10M$ (right). The potential is shown in a plane cutting through the center for the entire 160 × 160 Å domain along with a detail of the H3 tail region.

concentration the nonlinear behavior of the field extends farther into the solvent and limits the accuracy of *N*-body representation. In the limit of high salt concentration, the field varies over a smaller length scale, so the accuracy of the *N*-body field at a fixed resolution decreases as salt concentration increases.

At the lowest resolution and highest salt concentration examined ($N = 62$, $C_s = 0.10M$), the optimal *N*-body representation differs from the PBE field by an average of 31% in the region V'. The error for $N = 62$ is reduced to about 16% at $0.01M$ salt. The behavior is more favorable at higher resolution: error is less than 10% for $N = 277$ and 353 points over the full range of salt concentration.

The potentials predicted for the $N = 62$ and $N = 277$ approximations are compared to the electro-

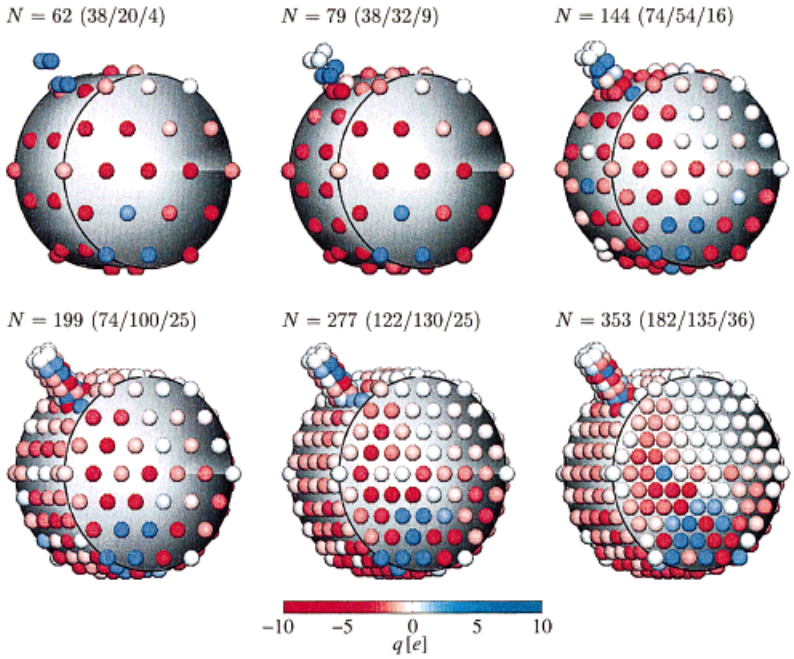

FIGURE 3 Distribution of charges on the particle surface S at six different resolutions ($N = 62$, 79, 144, 199, 277, and 353) for $C_s = 0.05M$. The distribution of charges over each region is reported in parenthesis by three numbers corresponding to (1) the top and bottom flat surfaces, (2) the circular outer edge, and (3) the H3 tail region of the nucleosome core model.

static potential associated with the PBE in the lower panels of Figure 2. Shown is a slice through the center of the computational domain, with the particle oriented as in Figure 1. We note that the potential is mostly negative around the edge, except for in the neighborhood of the H3 tail. Remarkably, even for the

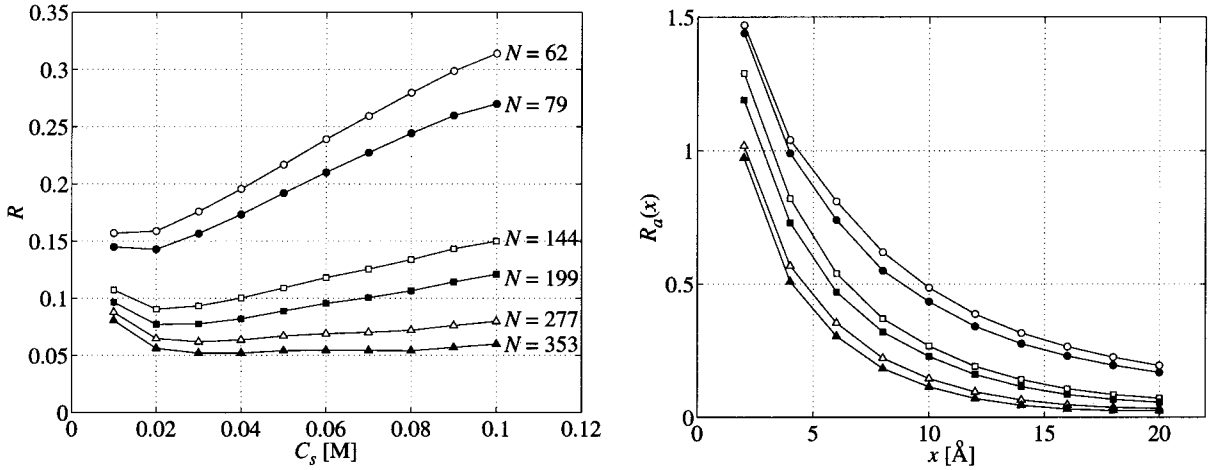

FIGURE 4 Assessment of accuracy of the N-body representation. Left panel: The relative error R associated with the N-body Debye–Hückel representation as a function of C_s (the bulk salt concentration in the solvent) for various charge set sizes N. Right panel: The relative error $R_a(\tilde{\mathbf{E}}, \mathbf{E}; x)$ associated with the N-body Debye–Hückel representation at a fixed distance x from the surface of the core particle as a function of x, for $C_s = 0.05M$. We calculate $R_a(x)$ by integrating the error over the surface $A(x)$: $R_a(\tilde{\mathbf{E}}, \mathbf{E}; x) = (1/S_x) \int_{A(x)} ((\|\mathbf{E} - \tilde{\mathbf{E}}\|)/\|\mathbf{E}\|)\, dA$, where $A(x)$ is defined as the surface outside the macromolecule that lies the normal distance x from the surface S. The normalizing factor S_x, is the area of $A(x)$. For the results shown here, the value $d = 10$ Å is used for evaluation R and optimizing the N charges.

62-charge case (where $R = 0.3$ for $C_s = 0.10M$), gross features of the PBE potential are reproduced with our discrete approximation. This low-order approximation is improved substantially as N is increased to 277, as shown in the figure.

We also note that the accuracy of the Debye–Hückel approximation increases with distance from the macromolecular surface. This behavior is evident in the right panel of Figure 4, which plots the error at a fixed distance from the surface $R_a(x)$ as a function of distance from the surface of the core particle for the salt concentration of $C_s = 0.05M$. $R_a(x)$ is measured as the average error over the points separated from the surface by distance x; see legend to Figure 4. At distances greater than the Debye length (13.5 Å for $C_s = 0.05M$), the error saturates to about 5% for $N = 277$ and $N = 353$.

The analyses presented here show that the 277-charge model for the nucleosome produces an accurate approximation to the PBE model. The much cheaper N-body approximation can be integrated easily into a general simulation protocol that requires evaluating salt-mediated electrostatic interactions at each step.

FUTURE APPLICATIONS

Our approach described here, DiSCO, for an N-body Debye-Hückel representation of a complex macromolecular system, was motivated by our effort to model the electrostatic forces acting on the chromatin fiber.[2] The basic building block of chromatin is illustrated in the upper panel of Figure 5: two-core particles connected by a linker DNA segment. The DNA is depicted as a red tube that wraps around the core particles and is continuous with the linker DNA segment. Previous models[34,35] for the structure and dynamics of chromatin have not considered the electrostatics of the highly charged core particle. This simplification has, to date, limited dynamic modeling of chromatin. By treating the flexible linker DNA using a polymer bead model[8–11] and exploiting the N-body charge model for the core, we can construct a polymer-level model for the chromatin fiber, as shown in the middle panel of Figure 5. This macroscale model (details presented in a forthcoming publication[2]) ignores the internal motions of atoms of the core particle and treats the core complex as a rigid body, as done for the several base pairs of linker DNA treated by each bead. With our approximate treatment of the electrostatics of this system, larger polynucleoseome systems can be modeled, such as the 48-nucleosome structure shown in the lower panel of Figure 5.

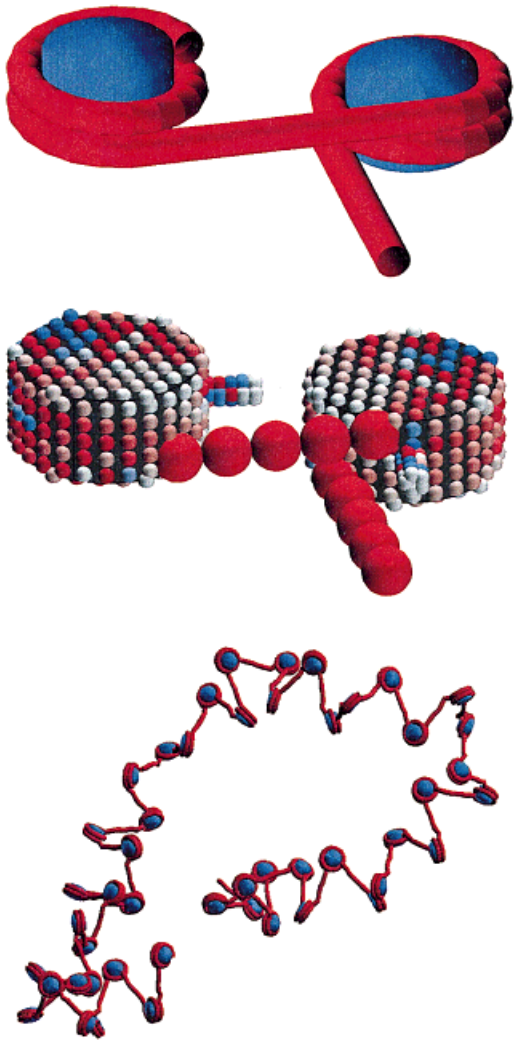

FIGURE 5 Construction of biophysical model for chromatin based on the N-body parameterization of core particle electrostatics. Upper panel: A dinucleosome is depicted as two core particle connected by linker DNA. DNA is represented as a continuous red tube. Middle panel: A macro-level model for the dincleosome is constructed using the N-body representation for core-particle effective charges and a bead-like model for the flexible linker DNA. Lower panel: A large polynucleosome model is constructed from a serial array of 48 nucleosomes. The configuration shown here was obtained from Monte Carlo sampling of the system at the monovalent salt concentration of 0.01M. Details of this model are presented in Ref. 2.

DiSCO requires both a PBE solver and a nonlinear minimization package. In practice, we have used the DelPhi package to obtain solutions to the PBE, but other software packages may be equally suitable. The quadratic convergence of our optimization package TNPACK[27–29] used here was crucial in making accurate minimizations of the residual function realizable.

The application of our methodology to a system as large and highly charged as the nucleosome core

particle demonstrates the potential for application of similar methods to other large systems. For other systems, the definition of the enclosing surface geometry and the distribution of charges on the surface will depend on the particular application. Surfaces may be constructed from ideal geometric shapes, as we have done here, or from automated routines for surface generation.

In general, DiSCO can be applied to a broad class of problems in which an efficient method for calculating the electrostatic interactions between biomolecules in solution is desired. The resulting potential is flexible in that it can be incorporated into most simulation protocols (energy minimization, Monte Carlo, MD, or BD). For example, the *N*-body approximation can help make PBE-based applications tractable (for a certain range of intermolecular distances), eliminating repeated PBE solutions as the relative orientation of the molecules changes. This approach can thereby help bridge details on the all-atom level with a macroscopic treatment of biopolymer systems. Interested users are invited to contact us for a program package.

The authors thank Dr. B. Honig and co-workers for providing the DelPhi and Grasp programs and Dr. J. A. McCammon for providing the UHBD package. We also thank Q. Zhang for calculations which contributed to Figure 4. Support by the National Science Foundation (ASC-9157582) and the National Institutes of Health (R01 GM55164) is gratefully acknowledged. T. Schlick is and investigator of the Howard Hughes Medical Institute.

REFERENCES

1. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Nature 1997, 389, 251–260.
2. Beard, D. A.; Schlick, T. Structure, submitted.
3. Izrailev, S.; Crofts, A. R.; Berry, E. A.; Schulten, K. Biophys J 1999, 77, 1753–1768.
4. Duan, Y.; Kollman, P. A. Science 1998, 282, 740–744.
5. Bashford, D.; Case, D. A. Ann Rev Phys Chem 2000, 51, in press.
6. Dominy, B. N.; Brooks, C. L., III. J Phys Chem 1999, 103, 3765–3773.
7. Fox, R. F. Phys Rep 1978, 48, 181–283.
8. Allison, S. A.; Austin, R.; Hogan, M. J Chem Phys 1989, 90, 3843–3854.
9. Chirico, G.; Langowski, J. Biopolymers 1994, 34, 415–433.
10. Schlick, T.; Li, B.; Olson, W. K. Biophys J 1994, 67, 2146–2166.
11. Jian, H.; Schlick, T.; Vologodskii, A. J Mol Biol 1998, 284, 287–296.
12. Ermak, D. L.; McCammon, J. A. J Chem Phys 1978, 69, 1352–1360.
13. Rotne, J.; Prager, S. J Chem Phys 1969, 50, 4831–4837.
14. Harvey, S. C. Proteins 1989, 5, 78–92.
15. Davis, M. E.; McCammon, J. A. Chem Rev 1990, 90, 509–521.
16. Honig, B.; Nicholls, A. Science 1995, 268, 1144–1149.
17. Elliot, R. S. Electromagnetics; McGraw-Hill: New York, 1966.
18. Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. Comp Phys Comm 1991, 62, 187–197.
19. Stigter, D. Biopolymers 1977, 16, 1435–1448.
20. Schlick, T.; Olson, W. K. J Mol Biol 1992, 223, 1089–1119.
21. Vologodskii, A. V.; Cozzarelli, N. R. Biopolymers 1995, 35, 286–296.
22. Gabdoulline, R. R.; Wade, R. C. J Phys Chem 1996, 100, 3868–3878.
23. Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K. A.; Honig, B. Proteins 1986, 1, 47–59.
24. Gilson, M. K.; Sharp, K. A.; Honig, B. J Comp Chem 1988, 9, 327–335.
25. Nicholls, A.; Honig, B. J Comp Chem 1991, 12, 435–445.
26. Honig, B.; Sharp, K.; Yang, A. J Phys Chem 1993, 97, 1101–1109.
27. Schlick, T.; Fogelson, A. ACM Trans Math Software 1992, 18, 46–70.
28. Schlick, T.; Fogelson, A. ACM Trans Math Software 1992, 18, 71–111.
29. Xie, D.; Schlick, T. ACM Trans Math Software 1999, 25, 108–122.
30. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeryer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179–5197.
31. Chin, K.; Sharp, K. A.; Honig, B.; Pyle, A. M. Nat Struct Biol 1999, 6, 1055–1061.
32. Gilson, M. K.; Honig, B. Proteins 1988, 4, 7–18.
33. Nicholls, A.; Sharp, K. A.; Honig, B. Proteins 1991, 11, 281–296.
34. Woodcock, C. L.; Grigoryev, S. A.; Horowitz, R. A.; Whitaker, N. Proc Natl Acad Sci USA 1993, 90, 9021–9025.
35. Ehrlich, L.; Munkel, C.; Chirico, G.; Langowski, J. CABIOS 1997, 13, 271–279.
36. Darden, T.; Perera, L.; Leping, L.; Pederson, L. Structure 1999, 7, R55–R60.