Constructing Irregular Surfaces to Enclose Macromolecular Complexes for Mesoscale Modeling Using the Discrete Surface Charge Optimization (DiSCO) Algorithm

QING ZHANG,[1] DANIEL A. BEARD,[2] TAMAR SCHLICK[1]
[1]*Department of Chemistry and Courant Institute of Mathematical Sciences, New York University and the Howard Hughes Medical Institute, 251 Mercer St., New York, New York 10012*
[2]*Department of Bioengineering, University of Washington, Box 352255, AERL 341, Seattle, Washington 98195*

Received 7 April 2003; Accepted 27 May 2003

Abstract: Salt-mediated electrostatics interactions play an essential role in biomolecular structures and dynamics. Because macromolecular systems modeled at atomic resolution contain thousands of solute atoms, the electrostatic computations constitute an expensive part of the force and energy calculations. Implicit solvent models are one way to simplify the model and associated calculations, but they are generally used in combination with standard atomic models for the solute. To approximate electrostatics interactions in models on the polymer level (e.g., supercoiled DNA) that are simulated over long times (e.g., milliseconds) using Brownian dynamics, Beard and Schlick have developed the DiSCO (Discrete Surface Charge Optimization) algorithm. DiSCO represents a macromolecular complex by a few hundred discrete charges on a surface enclosing the system modeled by the Debye-Hückel (screened Coulombic) approximation to the Poisson-Boltzmann equation, and treats the salt solution as continuum solvation. DiSCO can represent the nucleosome core particle (>12,000 atoms), for example, by 353 discrete surface charges distributed on the surfaces of a large disk for the nucleosome core particle and a slender cylinder for the histone tail; the charges are optimized with respect to the Poisson-Boltzmann solution for the electric field, yielding a ~5.5% residual. Because regular surfaces enclosing macromolecules are not sufficiently general and may be suboptimal for certain systems, we develop a general method to construct irregular models tailored to the geometry of macromolecules. We also compare charge optimization based on both the electric field and electrostatic potential refinement. Results indicate that irregular surfaces can lead to a more accurate approximation (lower residuals), and the refinement in terms of the electric field is more robust. We also show that surface smoothing for irregular models is important, that the charge optimization (by the TNPACK minimizer) is efficient and does not depend on the initial assigned values, and that the residual is acceptable when the distance to the model surface is close to, or larger than, the Debye length. We illustrate applications of DiSCO's model-building procedure to chromatin folding and supercoiled DNA bound to *Hin* and *Fis* proteins. DiSCO is generally applicable to other interesting macromolecular systems for which mesoscale models are appropriate, to yield a resolution between the all-atom representative and the polymer level.

 J Comput Chem 24: 2063–2074, 2003

Key words: electrostatics; nonlinear Poisson-Boltzmann equation; Debye-Hückel approximation; electric field refinement; electrostatic potential refinement; nucleosome core particle; discrete surface charges

Introduction

Biomolecular electrostatics governs many important processes associated with DNA and/or proteins.[1] However, evaluating the electrostatic component of the force in dynamic simulations is challenging because of the complex structures of macromolecules, the large number of solute atoms, and the huge amount of water molecules and salt ions. The third difficulty can be avoided by modeling the solvent as a continuum, a successful approach in many applications today. Several major types of implicit solvation models are widely used[2,3]: finite difference solutions of the nonlinear Poisson-Boltzmann equation (PBE),[4] boundary element methods (BEM),[5] and the generalized Born (GB) model.[6–10] The

Correspondence to: T. Schlick; e-mail: schlick@nyu.edu

Contract/grant sponsor: National Institutes of Health; contract/grant number: R01 GM55164

first approach solves the electrostatics potential at each point on a computational grid, given fixed charges on the macromolecules. The second approach can produce effective charges for the system, namely solute charges and induced surface charge, but its performance depends on the density of surface elements. In the third approach, the electrostatic potential is estimated based on effective Born radii and an analytical atom-pair energy function. Excellent packages are available for solving the PBE, such as that by the McCammon and Holst groups[11,12] and by Honig and coworkers[4,13] (see also Chapter 9 in ref. 14). When very large biopolymer systems that contain multiple proteins bound to long DNA are the targets for the long-time (millisecond and longer) studies, "mesoscale" type models are appropriate—which incorporate local details as necessary (like protein-bound sites) and approximations as possible (e.g., wormlike chain/bead model for supercoiled DNA of thousands of base pairs)—to make possible long simulations such as by Brownian dynamics.[15]

Beard and Schlick have recently developed the Discrete Surface Charge Optimization (DiSCO) algorithm[16] to address these difficulties with the motivation of studying chromatin folding.[17] Essentially, the nucleosome core with its wrapped DNA is modeled as a simple geometric object (a disk and a cylinder), based on the nucleosome core crystal structure.[18] A few hundred discrete surface charges[16] are assigned based on an optimization procedure that minimizes the Debye-Hückel (DH) approximation (electric field)[19,20] with respect to the nonlinear PBE. The linker DNA—which connects each nucleosome core particle to the next—can then be represented by the homogeneous wormlike chain/bead model developed by Allison and coworkers[21,22] and extended by Chirico and Langowski for supercoiled DNA[23,24] (see model developments for linear DNA,[25] circular DNA at low salt,[26] and circular DNA at high salt[27]). Thus, the DiSCO approach can reduce the electrostatic calculations for medium- and long-range macromolecular interactions and make possible large-scale rearrangements in Brownian dynamics simulations.[17] DiSCO has also been used more recently to study the *Hin*-mediated and *Fis*-enhanced DNA inversion reaction,[28,29] in which three protein-bound sites on long supercoiled DNA juxtapose to form a synaptic supercoiled complex.[30–32] Such applications can help in understanding systematic effects, such as chromatin folding as a function of salt, or the dependence of juxtaposition dynamics on the superhelical density and on the bound proteins.

An alternative algorithm termed ECM (Effective Charges for Macromolecules in solvent) developed by Gabdoulline and Wade[33] also uses effective charges (potential derived) to model macromolecules in implicit solvent, with the charges placed inside, rather than on the surface, of the macromolecular systems. ECM derives the effective charges by fitting them to reproduce the electrostatic potential calculated relative to the finite-difference solution of the linearized PBE. Because the approximate solution by linearizing PBE should satisfy $|e\phi| \ll 1$, where e is the elementary charge, this approach works well only when the electrostatic potential ϕ is small. As we will show here, a refinement based on the electric field ($\mathbf{E}$) rather than the electrostatic potential (ϕ, where $-\nabla\phi = \mathbf{E}$) is more robust. Furthermore, optimization in DiSCO using our truncated Newton minimizer TNPACK[34–36] is efficient and likely works better than the procedure used in ECM based on matrix inversion, which is computationally intensive and a potential source of error for large systems.

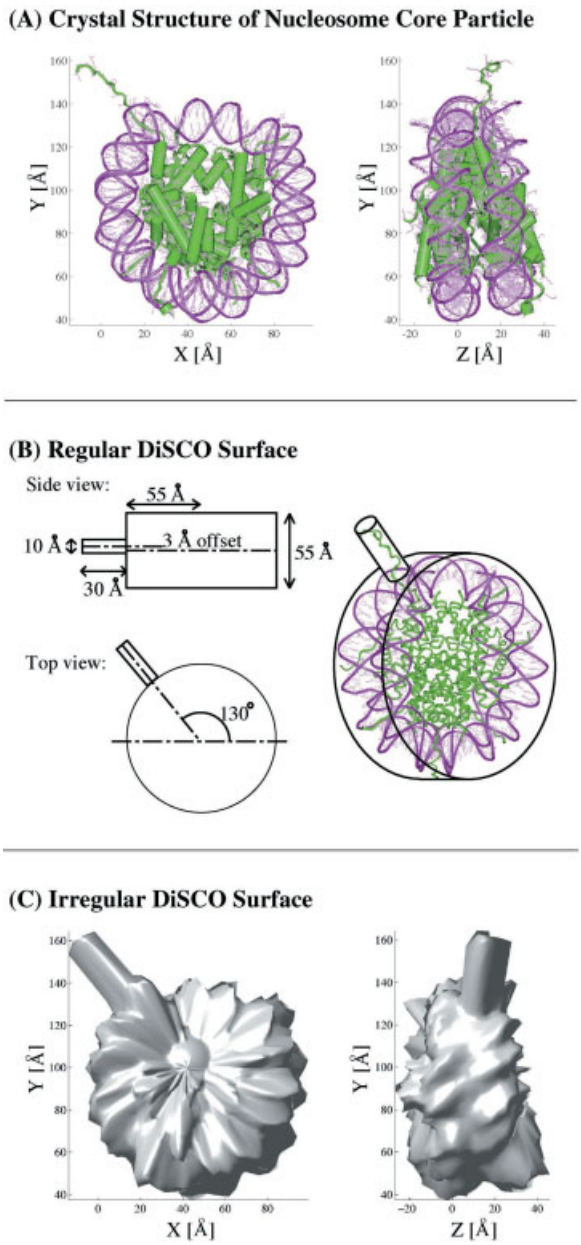

Figure 1. The crystal structure, regular DiSCO surface, and irregular DiSCO surface of the nucleosome core particle (PDB code: 1aoi).[18] [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

The first version of DiSCO was applied to model the nucleosome core particle (Fig. 1A),[18] with the following general protocol. After reading coordinates from the Protein Data Bank (PDB; www.pdb.org; nucleosome core particle, code 1aoi), DiSCO calculates the electrostatic potential ϕ by solving the nonlinear PBE using QNIFFT 1.2,[37–39] an outgrowth of the DelPhi program.[4,13] The corresponding electric field $\mathbf{E}$ is then calculated as $\mathbf{E} = -\nabla\phi$. For the nucleosome core particle, the surface is predefined—by an engineering-type construction—to be a composite of one large disk for the DNA-wrapped nucleosome core and a small cylinder

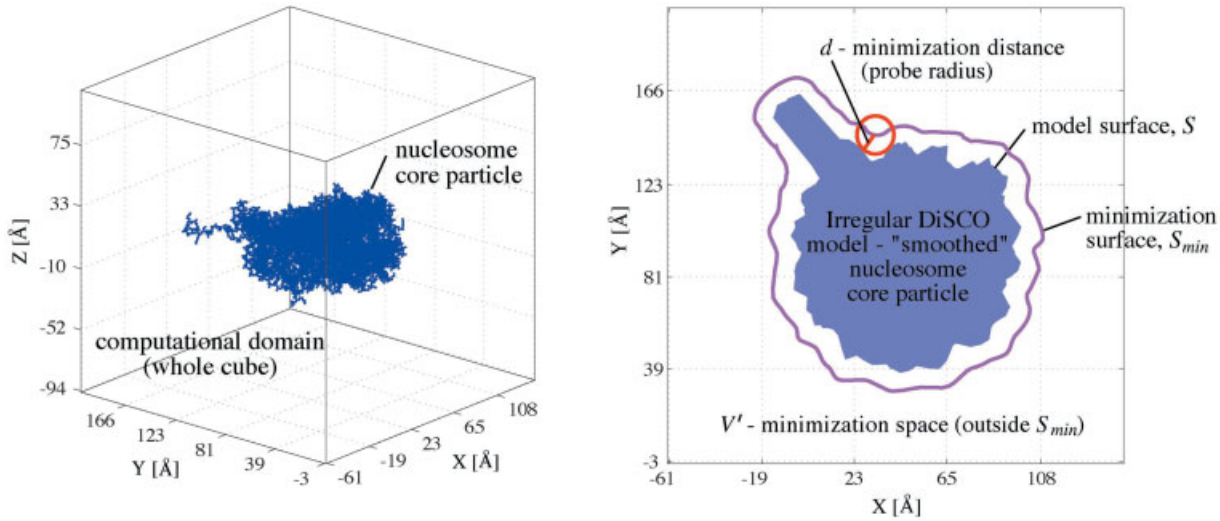

Figure 2. Left: the computational domain of the electrostatic potential and the electric field contains 65 × 65 × 65 grid points enclosing the nucleosome core particle. Right: the minimization space V' (of the residual) defines the computational domain separated from DiSCO surface by the minimization distance d. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

for the protruding H3 tail, as shown in Figure 1B. Initial values for the discrete charges are then assigned on the model surface,[16] after which DiSCO calculates the corresponding DH potential $\tilde{\phi}$ and associated electric field $\tilde{\mathbf{E}}$. The DH potential $\tilde{\phi}$ at position $\mathbf{r}_i$ is defined as

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

$$\kappa = \sqrt{\frac{8\pi e^2 N_A C_s}{k_B T \epsilon}} \tag{2}$$

where N is the number of discrete surface charges; q_j is the value of discrete surface charge j; r_{ij} is the distance between charge j and position i; k_B is the Boltzmann constant; T is the absolute temperature; ϵ is the dielectric constant; e is the elementary charge; N_A is the Avogadro constant; C_s is the monovalent salt concentration. (The DH potential is a linear homogeneous approximate solution to the nonlinear PBE if $|e\phi| \ll 1$.) DiSCO then uses the TNPACK minimizer developed by Schlick and coworkers[34–36] to optimize the values of the discrete surface charges q_j, $j = 1, 2, \ldots, N$, so as to minimize the residual R of the electric field $\mathbf{E}$ between the nonlinear PBE solution and the DH approximation. Finally, DiSCO outputs the optimal charges and the minimum residual.

Here, we report enhancements of the DiSCO package in terms of both modeling and computation. Namely, we now construct irregular, rather than regular surfaces, to enclose any general macromolecular system, and also allow the refinement (to produce the discrete surface charges) based on the electrostatic potential as an alternative to electric field-based refinement; we find the latter to be more robust. We evaluate performance of these enhancements by comparing the former models of the nucleosome core particle system with the new approximation, and also study the dependence of the residual R on the threshold distance d beyond which the DiSCO approximation is valid, and the dependence of the optimized discrete surface charges on their initial values. We also illustrate the DiSCO models for the recent *Hin*-mediated application.[28,29]

The article is organized as follows. In the next section, we describe how we construct irregular models enclosing general macromolecular systems and perform the optimization to assign discrete surface charges. We then describe results to evaluate $\mathbf{E}$ versus ϕ-based refinement, regular versus irregular models, effects of surface smoothing, and the performance of charge optimization.

Methods

Computing the Electrostatic Potential and Electric Field

DiSCO computes the electrostatic potential $\phi(\mathbf{r})$ by solving the nonlinear PBE for a monovalent salt solution using the QNIFFT 1.2 package.[37–39] DiSCO runs QNIFFT under the "coulombic" boundary condition followed by the focussing boundary condition, used to represent an infinite media. The computational domain is a 65 × 65 × 65-grid-point cube that encloses the whole system (e.g., nucleosome core particle), as shown in Figure 2 (left). The ratio between the largest dimension of the target macromolecule and the side length of the cube is 0.5 under the coulombic and 0.6 under the focussing boundary condition. This means that the side length of the cube is decreased under the focussing boundary condition in order to minimize the edge effect in the calculation under the coulombic boundary condition. The corresponding scales are 0.25 grid/Å and 0.30 grid/Å, respectively.

In solving the nonlinear PBE, atomic partial charges are assigned from the 1995 AMBER force field.[40] The atomic radii are

assigned with the extended atom radii based loosely on Mike Connolly's MS program.[41] The temperature T is set to 298 K. At this temperature, the solvent's dielectric constant ϵ is 78.3.[42] The dielectric constant inside the target macromolecule is assigned to a low value of 2 following the literature.[4,16] The salt concentration C_s is varied from 0.01 to 0.10 M, with charges optimized as a function of C_s. These assignments and settings echo those used in the original DiSCO package[16] except for the slight change in ε (78.3 now instead of 80).

After the electrostatic potential $\phi(\mathbf{r})$ is computed, the electric field $\mathbf{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r})$ is calculated, where differentiation is performed with respect to the independent variables $\mathbf{r}$ (Cartesian coordinates of grid points in the computational domain).

Building Surface Models

Programs such as GRASP[43] can also construct molecular surfaces of macromolecules. However, such surfaces require many more points to define than appropriate for long-time simulations of large systems. DiSCO aims to represent a macromolecular complex by a few hundred discrete surface charges.

The original models were built manually[16] and thus reflected regular domains, such as those constructed from cylinders and spheres. To apply DiSCO more generally, we developed the following algorithm to automatically construct geometric models for any macromolecular complex. There are four major steps in the irregular DiSCO model building procedure (Fig. 3), which we describe in turn.

Defining Surface Elements/Filling Empty Volumes/Initial Smoothing

Surface models are constructed in spherical coordinates (ρ, θ, φ), $\theta \in [0°, 180°]$, $\varphi \in [0°, 360°]$. The coordinate center of the target macromolecule is defined as the origin, with

$$\theta_j = (j-1) \times \Delta\theta, \quad j = 1, \ldots, j\_max,$$

$$\varphi_k = (k-1) \times \Delta\varphi, \quad k = 1, \ldots, k\_max,$$

$$\Delta\theta = \Delta\varphi = 4°, \quad j\_max = 180/\Delta\theta, \quad k\_max = 360/\Delta\varphi$$

Note that (j, k) indexes the volume element represented by $\theta_j : \theta_j + \Delta\theta$, $\varphi_k : \varphi_k + \Delta\varphi$, in which the algorithm finds an atom whose center is farthest away from the origin, and defines its distance vector as $\rho(j, k)$ (Pseudocode 1). The associated surface area $\rho(j, k)\sin(\theta)\Delta\theta\Delta\varphi$ is defined as the surface element $SE(j, k)$, as shown in Figure 3A (left):

```
/* Pseudocode 1: Defining Surface Elements */
/* Natom: total number of atoms in the target macromolecule */
/* Initialize all ρ(j, k) to zero */
for i = 1, Natom
  (x_i, y_i, z_i) ⇒ (j, k); /* find the volume element for atom i
  */
  ρ_i = √(x_i^2 + y_i^2 + z_i^2);
  if ρ(j, k) < ρ_i
    ρ(j, k) = ρ_i;
  end
end
```

Because not every 3D slice contains atoms, that is, $\rho(j, k)$ may still be 0 for some (j, k) after Pseudocode 1, we thus implement Pseudocode 2 to perform the following operations. If $\rho(j, k) = 0$, then $\rho(j, k)$ will be assigned to the average value of its nearest nonzero neighbors $[\rho(j - nb, k), \rho(j + nb, k), \rho(j, k - nb), \rho(j, k + nb)]$, where nb is increased from 1 to 3 until a nonzero neighbor is found. For certain $\rho(j, k)$, no nonzero neighbors can be found even if nb has reached 3. Then the filling procedure continues until all $\rho(j, k)$ are nonzero. We denote this process as "filling empty volumes," as summarized below:

```
/* Pseudocode 2: Filling Empty Volumes */
/* Nzero: total number of ρ(j, k) whose value is 0 */
Nzero = 0;
for j = 1, j_max
  for k = 1, k_max
    if ρ(j, k) == 0
      Nzero = Nzero + 1;
    end
  end
end
while Nzero > 0
  for j = 1, j_max
    for k = 1, k_max
      if ρ(j, k) == 0
        for nb = 1, 3
          if ρ(j − nb, k) > 0 or ρ(j + nb, k) > 0
          or ρ(j, k − nb) > 0 or ρ(j, k + nb) > 0 /* 4
          neighbors */
            ρ(j, k) = (sum of the nonzero neighbors) / (number of the nonzero neighbors);
            Nzero = Nzero − 1;
            break; /* exit current loop of nb */
          end
        end
      end
    end
  end
end
```

To smooth the model surface, the program executes "Initial Smoothing" (rounds of smoothing, $RS = 1$). If $\rho(j, k)$ is smaller than all of its four nearest neighbors $[\rho(j - 1, k), \rho(j + 1, k), \rho(j, k - 1), \rho(j, k + 1)]$, then $\rho(j, k)$ is assigned to the average value of these neighbors. This process (Pseudocode 3) is only executed once over $\rho(j, k)$ in this step. The target macromolecule now has a closed model surface enclosing the macromolecule itself (Fig. 3A, right).

```
/* Pseudocode 3: Smoothing */
/* Nmini: total number of ρ(j, k) whose value is smaller than
any of its 4 near neighbors */
/* RS_MAX: maximal rounds of smoothing. For 1 round of
smoothing, RS_MAX = 1 */
Nmini = 1; RS = 0;
while Nmini > 0 and RS <= RS_MAX
  Nmini = 0; RS = RS + 1;
  for j = 1, j_max
```

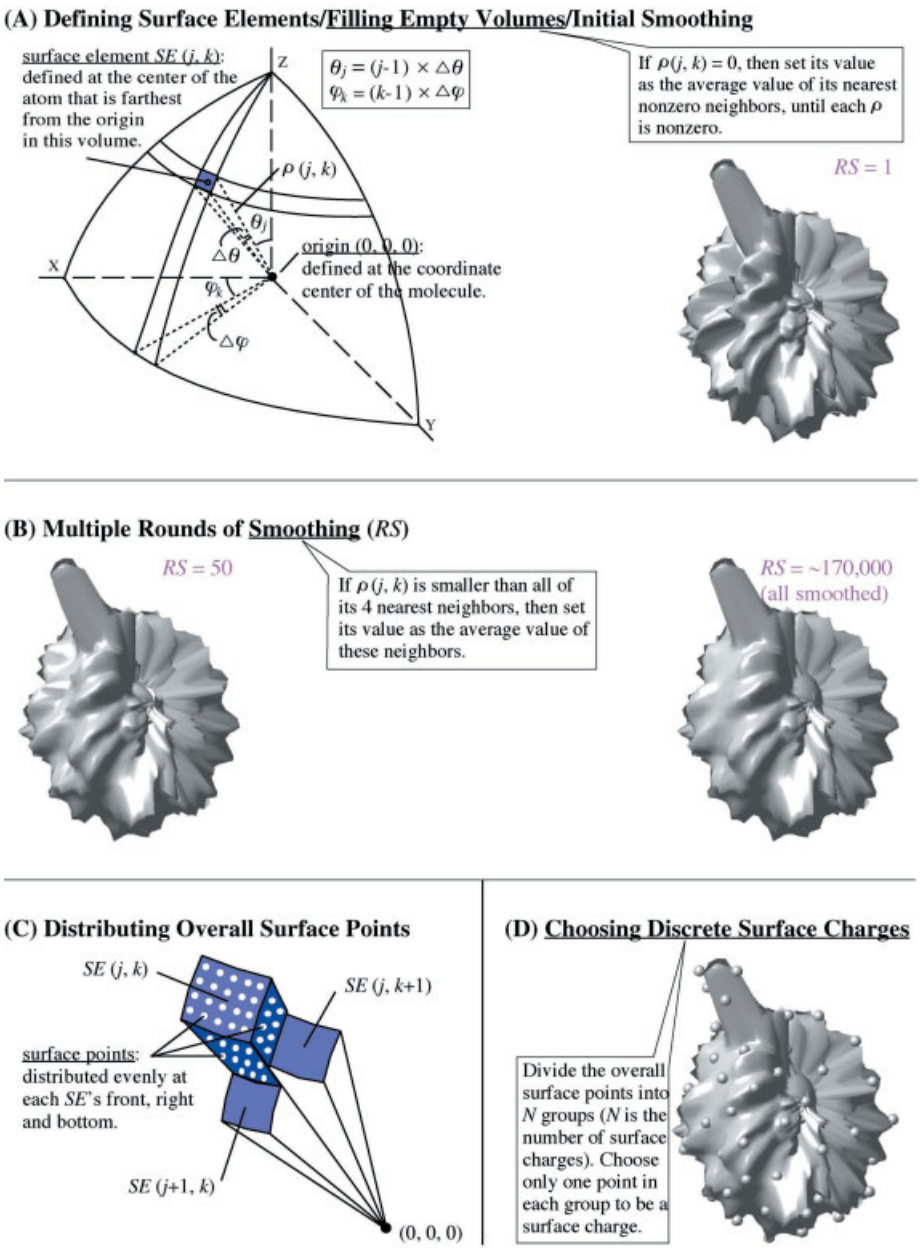

Figure 3. Irregular DiSCO model building procedure. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

```
    for k = 1, k_max
      if ρ(j, k) < ρ(j − 1, k) and ρ(j, k) < ρ(j + 1, k)
      and ρ(j, k) < ρ(j, k − 1) and ρ(j, k) < ρ(j, k + 1)
      /* 4 neighbors */
        ρ(j, k) = (sum of all the 4 neighbors) / 4;
        Nmini = Nmini + 1;
      end
    end
  end
end
```

Multiple RS (RS > 1)

To make the model surface smoother and simpler, we execute multiple *RS*. Figure 3B shows two smoothed model surfaces as obtained in the rounds $RS = 50$ and $RS = \sim 170{,}000$. The latter is the irregular model surface used for the nucleosome core particle (Fig. 1C).

Distributing Overall Surface Points

Because the $\rho(j, k)$ values are reassigned during the *RS*, the corresponding surface elements $SE(j, k)$ are redefined automati-

Table 1. Debye Length λ_D at Different Salt Concentrations C_s ($T = 298$ K).

C_s [M]	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
λ_D [Å]	30.4	21.5	17.6	15.2	13.6	12.4	11.5	10.8	10.1	9.6

cally. Because each surface element $SE(j, k)$ usually has a different value of $\rho(j, k)$ from its neighbors, the program evenly distributes surface points not only on the front face of surface element $SE(j, k)$, but also on its right side and bottom (Fig. 3C). These two surfaces join the surface element $SE(j, k)$ to its right-side neighbor $SE(j, k + 1)$ and bottom-side neighbor $SE(j + 1, k)$, respectively. [For the left and top side, the program has already distributed surface points when it operates on $SE(j, k - 1)$ and $SE(j - 1, k)$, respectively.] Thus, the resulting surface points are evenly distributed on the whole model surface. The density of the surface points is set to 4.0 per Å^2 in the program.

Choosing Discrete Surface Charges

The set of N discrete surface charge points is selected as a subset of the set of previously generated surface points. To distribute surface charge points evenly, the algorithm divides the set of surface points into N groups. In each group the surface charge position is assigned as the point farthest away from the the surface charge points already assigned in other groups (Fig. 3D). (The first chosen surface charge point can be any surface point in the first group.) In our nucleosome model, we consider six N values: 62, 79, 144, 199, 277, 353 (to be comparable with the regular DiSCO models[16]).

The entire DiSCO model-building procedure (with ~170,000 *RS*) only takes 47 s of CPU time on the SGI R12000 for the nucleosome core particle (12,386 total atoms for the proteins and DNA).

Optimizing Discrete Surface Charges

The initial values of all the discrete surface charges $q(j)$, $j = 1, 2, \ldots, N$ are assigned to zero. (We show that this assignment is reasonable in the Results section.) DiSCO uses our group's TNPACK[34–36] (a truncated Newton minimization package) to minimize the residual of the electric field (or the electrostatic potential) by optimizing the discrete surface charges. The formulae for the residuals are the following:

$$R_\phi = \frac{1}{N_v} \sum_{i=1}^{N_v} \frac{|\tilde{\phi}(i) - \phi(i)|}{|\phi(i)|} \tag{3}$$

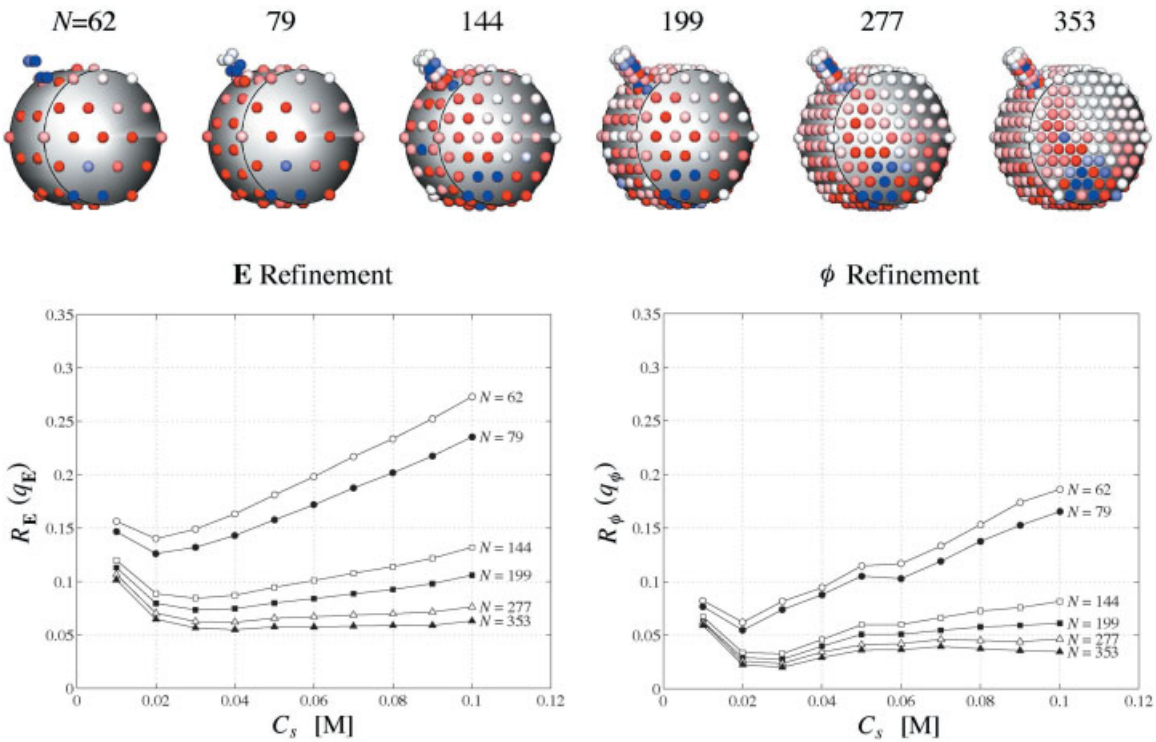

Figure 4. Top: the original regular surface models[16] of the nucleosome core particle with different numbers of discrete surface charges N. Bottom: the residual R as a function of the salt concentration C_s for different N, as shown on top, with $d = 10$ Å (the minimization distance) based on the **E** refinement (left) and ϕ refinement (right). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

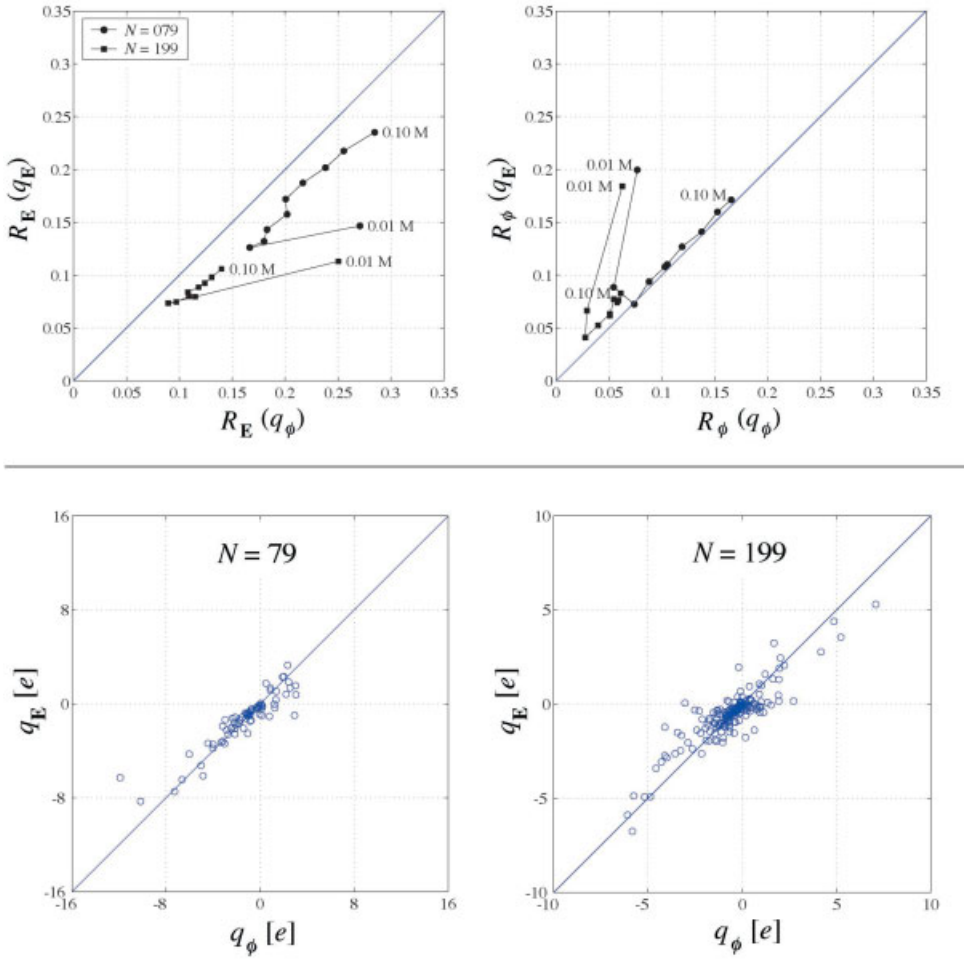

Figure 5. Top: comparison of the residual of electric field $R_{\mathbf{E}}$ (left) and the residual of electrostatic potential R_ϕ (right) between **E** refinement and ϕ refinement. Bottom: comparison between the optimized discrete surface charge sets $q_{\mathbf{E}}$ (from the **E** refinement) and q_ϕ (from the ϕ refinement) with $d = 10$ Å and $C_s = 0.05$ M, for $N = 79$ (left) and $N = 199$ (right). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

or

$$R_{\mathbf{E}} = \frac{1}{N_v}\sum_{i=1}^{N_v}\frac{|\tilde{\mathbf{E}}(i) - \mathbf{E}(i)|}{|\mathbf{E}(i)|} \tag{4}$$

where R_ϕ, $R_{\mathbf{E}}$ are the residuals of the electrostatic potential and electric field, respectively; ϕ is the electrostatic potential (of the target macromolecule) solved from the nonlinear PBE; $\tilde{\phi}$ is the DH potential of the model; **E** and $\tilde{\mathbf{E}}$ are the corresponding electric fields; and N_v is the number of grid points in the minimization space V' (Fig. 2, right).

The minimization of the residuals only operates on the grid points in the computational domain V' outside the minimization surface S_{min} (Fig. 2). The surface S_{min} is separated from the model surface S by a distance d (Fig. 2, right). To satisfy the condition of the DH approximation ($|e\phi| \ll 1$), the distance d is usually set to a value greater than the Debye length[16]:

$$\lambda_D = \frac{1}{\kappa} \tag{5}$$

where κ is defined in eq. (2). The Debye length values at different salt concentrations are listed in Table 1.

The original DiSCO algorithm optimized the discrete surface charges in terms of the electric field. The ECM algorithm[33] derives effective charges in terms of the electrostatic potential. To test which refinement (electric field or electrostatic potential) is more robust, we have modified the DiSCO algorithm to optimize the discrete surface charges based on the electrostatic potential. Because we find the latter less robust, we retain the original procedure in the distributed package.

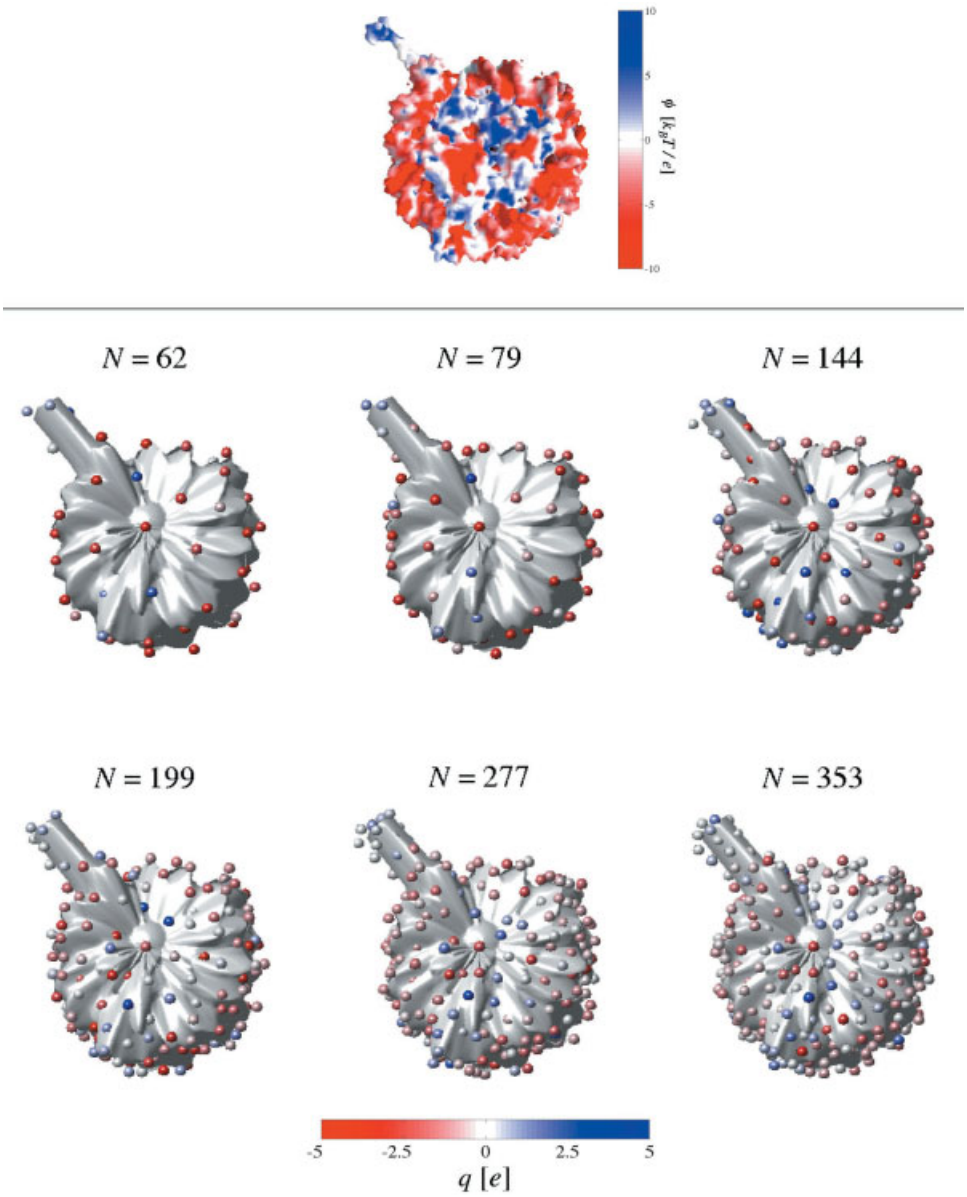

Figure 6. Top: the electrostatic potential on the molecular surface of the nucleosome core particle, visualized by GRASP, at C_s = 0.05 M. The electrostatic potential is computed by QNIFFT. Bottom: irregular DiSCO surface models with different numbers of discrete surface charges N, optimized with d = 10 Å and C_s = 0.05 M. The color of a charge represents it value according to the scale at bottom. The models are visualized with the program VIEWMODEL.[44] Each charge is visualized with a 2.5 Å radius, and each $\rho(j, k)$ (see Fig. 3) has been reduced by 2 Å to make the charges clear.

Results

Electrostatic Potential Refinement Versus Electric Field Refinement

To compare the potential-based (ϕ) refinement with the electric field-based (**E**) refinement, we use the original regular surface models[16] of the nucleosome core particle, as shown in Figure 1B. We define the resulting optimized charge sets as (q_ϕ) and ($q_{\mathbf{E}}$), respectively. The residual R as a function of the salt concentration C_s is calculated for different numbers of discrete surface charges N, varying from 62 to 353, as shown at the top of Figure 4 based on both refinements.

From Figure 4 we see that the residual of **E** ($R_{\mathbf{E}}$) is about twice as large as the residual of ϕ (R_ϕ). This is reasonable given the relationship between **E** and ϕ. Namely, when QNIFFT solves the nonlinear PBE using finite differences,[13] every grid point (except those on the boundary) is surrounded by six nearest grid points; because $\mathbf{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r})$, one error of ϕ at one grid point causes

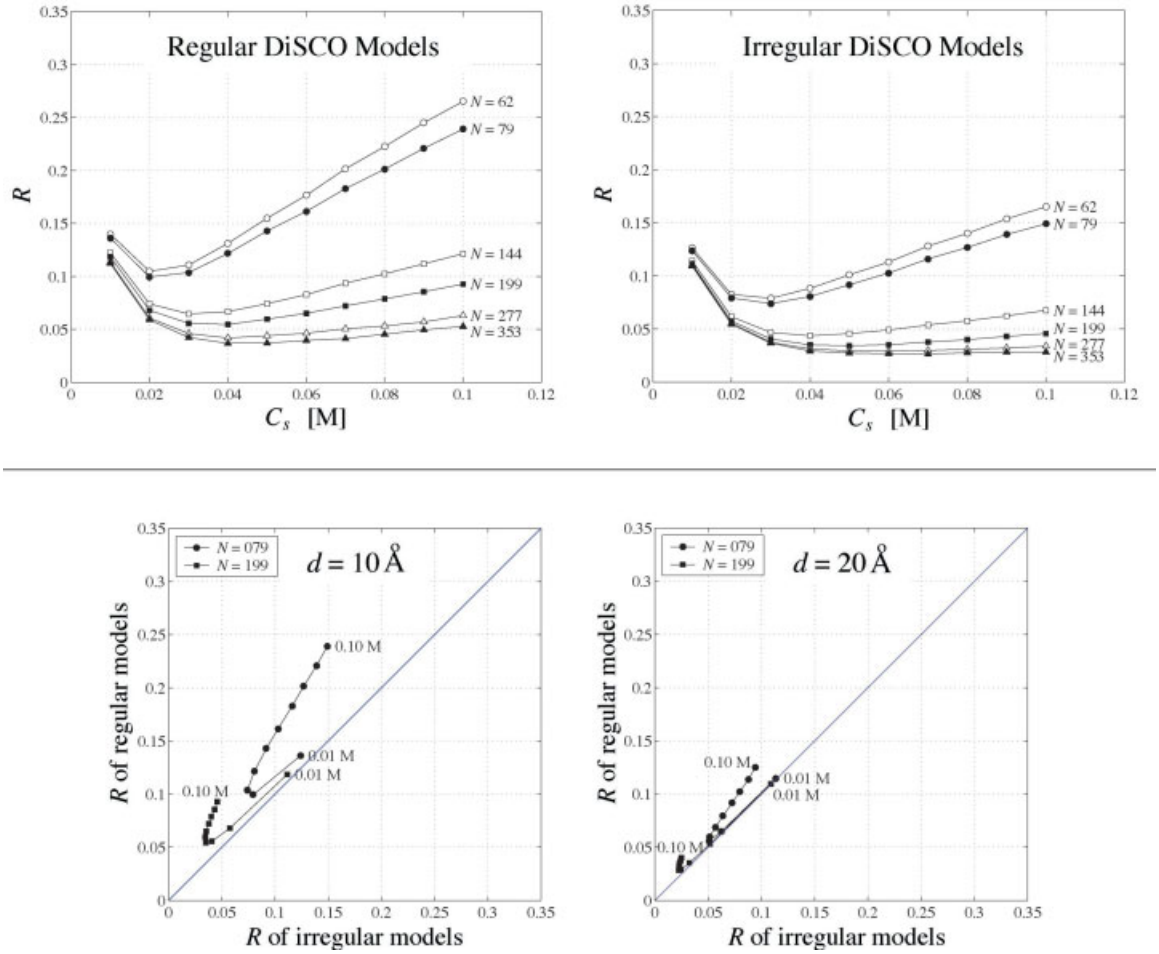

Figure 7. Top: the residual R as a function of the salt concentration C_s for different numbers of discrete surface charges N with $d = 10$ Å for regular (left) versus irregular (right) DiSCO surfaces. Bottom: comparison of the residual R between regular versus irregular DiSCO models at $d = 10$ Å (left) and $d = 20$ Å (right). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

errors of **E** at its six nearest grid points, each of which only has an error in one of its three components (E_x, E_y, E_z). Thus, one unit error of ϕ can cause about $6 \times 1/3 = 2$ corresponding units of error for **E**. We also note that the residual R in both cases increases at the low end and the high end of the salt concentrations. This may be caused by the nonlinear electrostatic effects on one hand and numerical errors on the other, related to our representation of fewer DH charges than actual molecular partial charges and the inhomogeneous medium of our system.

We also compute the residual of **E** ($R_{\mathbf{E}}$) from the ϕ refinement and the residual of ϕ (R_ϕ) from the **E** refinement. We compare $R_{\mathbf{E}}$ from both refinements in Figure 5 (top left), and R_ϕ from both refinements in Figure 5 (top right). The ϕ refinement gives 2–4% higher residuals $R_{\mathbf{E}}$ than the **E** refinement, while the **E** refinement gives only 0–1% higher R_ϕ than the ϕ refinement, for all salt concentrations except 0.01–0.02 M.

These tests show that the **E** refinement is much more robust than the ϕ refinement. The reason may be that all three components of **E** (E_x, E_y, E_z) are considered in the **E** refinement, while ϕ is only a scalar in the ϕ refinement.

We also compare the actual values of the optimized discrete surface charge sets $q_{\mathbf{E}}$ and q_ϕ with $d = 10$ Å and $C_s = 0.05$ M in Figure 5 (bottom) for $N = 79$ and $N = 199$. The figure clearly indicates that both refinements produce very similar and comparable optimal discrete surface charges.

Choosing the refinement type is system dependent. However, over a wide range of salt concentrations (≥0.02 M), the **E** refinement can yield a much more robust representation of macromolecules. Therefore, all the following work is based on the **E** refinement, and R denotes the residual of **E** based on the **E** refinement [$R = R_{\mathbf{E}}(q_{\mathbf{E}})$].

Regular Versus Irregular Models

The previous regular DiSCO models[16] for the nucleosome core particle had an ideal shape: a disk and a small cylinder (Fig. 1B). In comparison, Figure 6 (bottom) shows the models produced with our new irregular model building procedure (Fig. 3) for $N = 62$, 79, 144, 199, 277, 353, with $d = 10$ Å and $C_s = 0.05$ M. These charges have been optimized with TNPACK. The color of a charge represents its electronic charge, with blue representing positive values and red

Table 2. Comparison of the Residual R for Different Numbers of Discrete Surface Charges N, with $d = 10$ Å and $C_s = 0.05$ M, for Five Model Surfaces Listed in the Order of Simple to Complex.

N	Regular surface	Irregular surface			GRASP model
		All smoothed	$RS = 50$	$RS = 1$	
199	5.96%	3.43%	3.86%	4.05%	4.28%
353	3.70%	2.75%	2.79%	3.03%	3.37%
500	—	2.52%	2.60%	2.60%	3.13%
1000	—	2.41%	2.38%	2.42%	2.52%
2000	—	2.33%	2.30%	2.38%	2.34%

For each N, the underlined R indicates the smallest residual.

representing negative values (see scale at bottom). The dielectric constant ϵ inside and outside the model surface has the same value as that of the solvent, and the surface charges have no size. The models are visualized by our program VIEWMODEL.[44]

Figure 6 (top) shows the electrostatic potential on the molecular surface of the nucleosome core particle built on the crystallographic coordinates[18] as visualized by GRASP[43] ($C_s = 0.05$ M). The electrostatic potential is computed by QNIFFT. The molecular surface is also built by GRASP (using a reinvented "Marching Cubes" algorithm to produce a surface tessellation). We can see that the distribution of the discrete surface charges is consistent with the molecular surface potential map. (Note the positively charged histone tails and the negatively charged nucleosome exterior where DNA wraps around.)

The residual R as a function of the salt concentration C_s for different numbers of discrete surface charges N with $d = 10$ Å is shown in Figure 7 (top). Clearly the new irregular surface models reduce the residual R significantly. Still, while a regular surface could be constructed for the nucleosome, this is not possible in general, and the new approach can automatically be applied to any macromolecular complex. Note that the slight difference we see in R between Figure 4 (bottom left) and Figure 7 (top left) results mainly from our lower setting in the latter case for the QNIFFT convergence parameter (10^{-8} instead of 10^{-3}).

In Figure 7 we also compare the residual R for $N = 79$ and $N = 199$ at two d values 10 Å (bottom left) and 20 Å (bottom right). We see that, in both cases, the irregular surfaces yield smaller residuals. At the larger distances, residuals are smaller for both models because the DH approximation is more accurate the farther we are from the surface. Even with $d = 20$ Å, the irregular models yield a better representation of the nucleosome core particle, decreasing R by 0–3.5% for $N = 79$, and by 0–2% for $N = 199$.

Effects of Surface Smoothing

The irregular DiSCO models can reduce the residual significantly, because their surface is closer to the molecular surface. However, the molecular surface is clearly complex and requires sufficient smoothing to be practical. Table 2 shows the residual R for different numbers of discrete surface charges N with $d = 10$ Å and $C_s = 0.05$ M, from five kinds of model surfaces in order of decreasing smoothing, from left to right: original DiSCO model (Fig. 1B), all smoothed (Fig. 1C and 3B, right), $RS = 50$ (Fig. 3B, left), $RS = 1$ (Fig. 3A, right), and GRASP model (Fig. 6, top). The discrete surface charge assignment of the GRASP model is the same as that of our models (Fig. 3D). In Table 2, each underlined R indicates the smallest residual for the corresponding number of discrete surface charges N. We see that for $N = 199$, 353, and 500, the "all smoothed" DiSCO model surface obtains the smallest residual R (3.43, 2.75, and 2.52%, respectively); for $N = 1000$ and 2000, the $RS = 50$ model surface obtains the smallest R (2.38, 2.30%, respectively). Clearly, the greater the number of surface charges, the better the representation of a complex macromolecule. The GRASP model surface is too complex for practical simulations of long time.

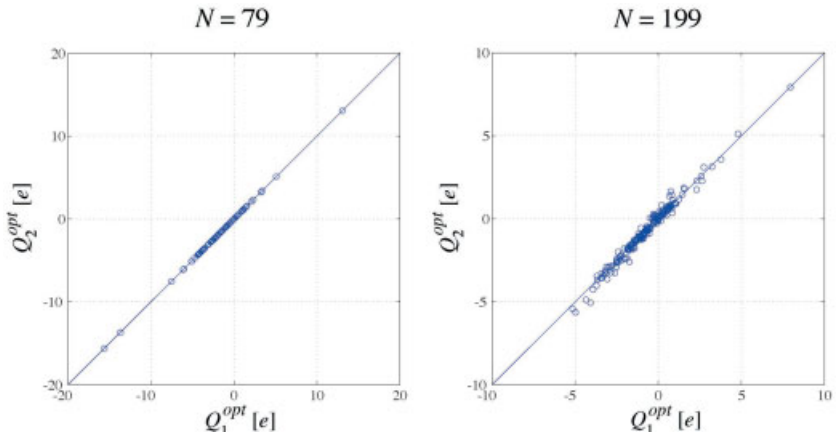

Figure 8. Comparison between the sets of optimal discrete surface charges Q_1^{opt} and Q_2^{opt} corresponding to two sets of initial values Q_1 (0, 0, . . .) and Q_2 (*rand*, *rand*, . . .) for $N = 79$ (left) and $N = 199$ (right), where rand is a random number between 0 and 1. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

Table 3. CPU Time t Used by TNPACK to Minimize the Residual of the Electric Field on SGI R12000, for Different Numbers of Discrete Surface Charges N with $d = 10$ Å and $C_s = 0.05$ M.

N	62	79	144	199	277	353	500	1000	2000
t [min]	0.47	0.38	2.30	3.09	4.42	5.60	8.12	15.36	33.63

The number of grid points in the minimization space V' (Fig. 2, right) is $N_v = 1084$.

Dependence of Charge Optimization on Initial Values and CPU Time

In our original algorithm, we have set the initial values of the discrete surface charges according to the surface charge density $\sigma(\mathbf{r})$ computed from the electric field $\mathbf{E}(\mathbf{r})$.[16] We now prove that this step is not necessary. The minimization program TNPACK[34–36] used for DiSCO can find the optimal charges just as well when the initial values are arbitrarily but reasonably assigned.

We experiment with two sets of initial values: $Q_1 = 0, 0, \ldots$; $Q_2 = rand, rand, \ldots$, where rand is a random number between 0 and 1. Q_1 is the set of initial values used above that yielded a distribution of optimal charges (Fig. 6, bottom) consistent with the molecular surface potential map (Fig. 6, top). The corresponding sets of optimal charges Q_1^{opt} and Q_2^{opt} are compared in Figure 8 (left for $N = 79$ and right for $N = 199$). We see that DiSCO finds nearly the same set of optimal charges, given either Q_1 or Q_2 for $N = 79$ and $N = 199$.

TNPACK not only finds optimal charges given reasonable initial values, but also performs the optimization very efficiently. Table 3 shows the CPU time t used by TNPACK to minimize the residual of the electric field on SGI R12000 as a function of increasing numbers of discrete surface charges N with $d = 10$ Å and $C_s = 0.05$ M. The number of grid points $N_v = 1084$ in the minimization space V' (Fig. 2, right).

Discussion

The nucleosome core particle, which has more than 12,000 atoms, now can be represented as an irregular DiSCO surface of 353 discrete charges with only a ~3% residual of the electric field over a wide range of salt concentration ($C_s \geq 0.04$ M). As another application, we illustrate in Figure 9 the irregular surface construction for the *Hin*/*Fis* DNA inversion system.[28,29] The computational requirements in both cases are reduced significantly for the medium- and long-range macromolecular interactions and allow long-time Brownian dynamics simulations to study interesting biological phenomena.[17,28,29]

We have shown that the **E** refinement is much more robust, likely because the three components of the electric field (E_x, E_y, E_z) are optimized simultaneously; while in the ϕ refinement the electrostatic potential ϕ is only scalar. For molecular dynamics applications, the **E** refinement is preferable because the electric field directly produces the electrostatic force. In computing free

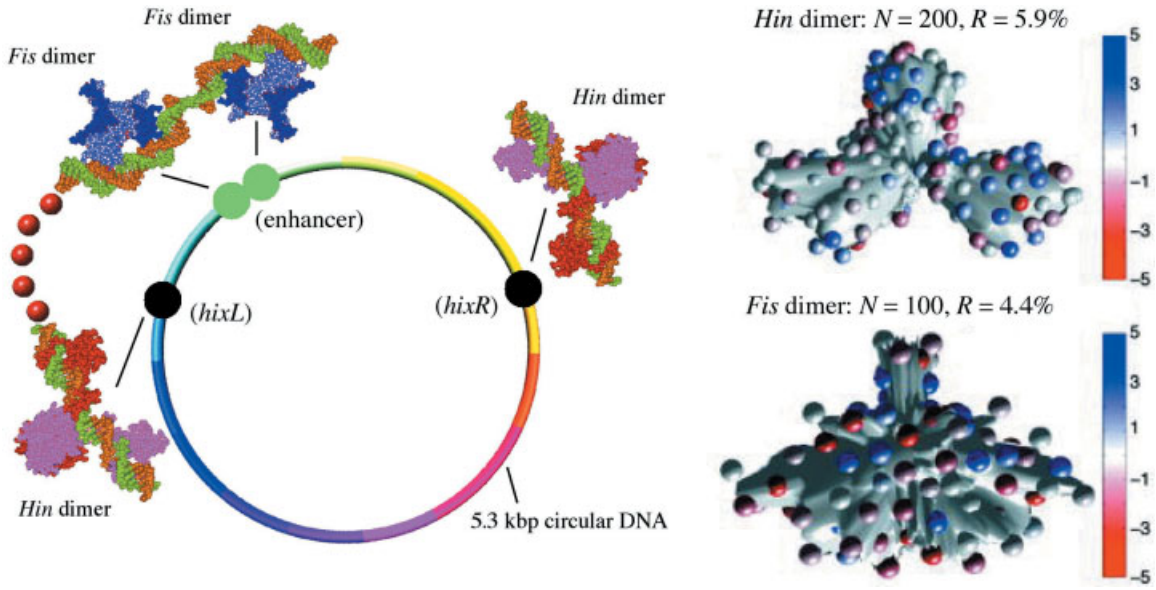

Figure 9. DiSCO models for supercoiled DNA with bound proteins in the *Hin*-mediated inversion system (left). DiSCO represents a *Hin* dimer as 200 discrete surface charges with a 5.9% residual (right top) and a *Fis* dimer as 100 charges with a 4.4% residual (right bottom). The *Hin* dimer and the *Fis* dimer have more than 8,200 and 3,600 atoms, respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

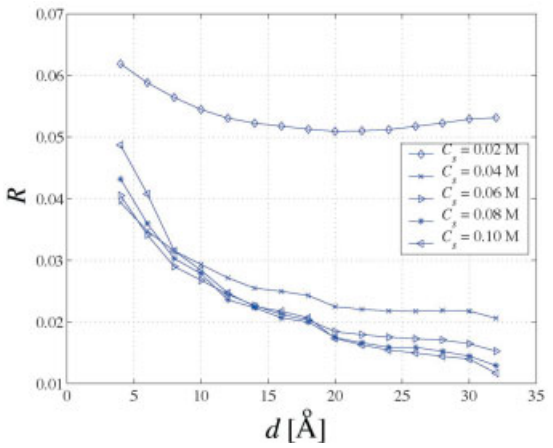

Figure 10. The residual R as a function of the minimization distance d at different salt concentrations C_s for $N = 353$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

energies, both refinements can be used as both perform well in terms of reproducing the electrostatic potential.

The algorithm developed here for building irregular surface models is simple and can be applied to any complex macromolecular system. As mesoscale models become important for bridging all-atom resolution with long-time polymer behavior, the DiSCO approach will be useful.

Note that the DH surface approach is only appropriate for studying the medium- and long-range macromolecular interactions and not short-range interactions or binding events, because the residual R is acceptable only when sufficiently far from the surface (i.e., d is close to, or greater than, the Debye length λ_D). Figure 10 shows the residual R as a function of the minimization distance d at different salt concentrations C_s for $N = 353$. The residual R decreases as d increases. The Debye length λ_D is shown in Table 1. For short-range interactions, a new short-range potential that is more accurate than the DH approximation must be developed.

DiSCO models are built based on the static geometry of macromolecules. For macromolecules with flexible conformations, we might compute DiSCO surfaces as a function of conformational changes, or DiSCO surfaces only for the rigid parts of macromolecules. The latter approach has been applied in refs. 17, 28, and 29.

We invite users to apply our DiSCO program, which can be downloaded from http://monod.biomath.nyu.edu/index/software/DiSCO. The VIEWMODEL program (visualizing DiSCO models) can be downloaded from http://monod.biomath.nyu.edu/~qzhang/Softwares.htm.

Acknowledgments

The authors thank Dr. Barry Honig and coworkers for providing the GRASP programs, and Dr. Kim Sharp and coworkers for developing the DelPhi program QNIFFT. We also appreciate discussions about TNPACK with Dr. Dexuan Xie, discussions about AMBER force field with Dr. Xiaoliang Qian, help from Dr. Daniel Strahs, and communications with Dr. Jing Huang. Tamar Schlick is an investigator of the Howard Hughes Medical Institute.

References

1. Fogolari, F.; Brigo, A.; Molinari, H. J Mol Recognit 2002, 15, 377.
2. Cramer, C. J.; Truhlar, D. G. Chem Rev 1999, 99, 2161.
3. Davis, M. E.; McCammon, J. A. Chem Rev 1990, 90, 509.
4. Honig, B.; Nicholls, A. Science 1995, 268, 1144.
5. Zauhar, R. J.; Morgan, R. S. J Comput Chem 1988, 9, 171.
6. Cramer, C. J.; Truhlar, D. G. J Am Chem Soc 1991, 113, 8305.
7. Cramer, C. J.; Truhlar, D. G. Science 1992, 256, 213.
8. Bashford, D.; Case, D. A. Annu Rev Phys Chem 2000, 51, 129.
9. Tsui, V.; Case, D. A. Biopolymers 2000, 56, 275.
10. Onufriev, A.; Case, D. A.; Bashford, D. J Comput Chem 2002, 23, 1297.
11. Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. Comp Phys Commun 1991, 62, 187.
12. Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. Comp Phys Commun 1995, 91, 57.
13. Nicholls, A.; Honig, B. J Comput Chem 1991, 12, 435.
14. Schlick, T. Molecular Modeling and Simulation: An Interdisciplinary Guide; Springer-Verlag: New York, 2002, 298–304. (http://www.springer-ny.com/detail.tpl?isbn=038795404X)
15. Schlick, T.; Beard, D. A.; Huang, J.; Strahs, D.; Qian, X. IEEE Comput Sci Eng 2000, 2, 38.
16. Beard, D. A.; Schlick, T. Biopolymers 2001, 58, 106.
17. Beard, D. A.; Schlick, T. Struct Fold Des 2001, 9, 105.
18. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Nature 1997, 389, 251.
19. Stigter, D. Biopolymers 1977, 16, 1435.
20. Schlick, T.; Li, B.; Olson, W. K. Biophys J 1994, 67, 2146.
21. Allison, S. A. Macromolecules 1986, 19, 118.
22. Allison, S. A.; Austin, R.; Hogan, M. J Chem Phys 1989, 90, 3843.
23. Chirico, G.; Langowski, J. Biopolymers 1994, 34, 415.
24. Chirico, G.; Langowski, J. Biophys J 1996, 71, 955.
25. Jian, H.; Vologodskii, A. V.; Schlick, T. J Comp Phys 1997, 136, 168.
26. Jian, H.; Schlick, T.; Vologodskii, A. J Mol Biol 1998, 284, 287.
27. Huang, J.; Schlick, T.; Vologodskii, A. Proc Natl Acad Sci USA 2001, 98, 968.
28. Huang, J.; Schlick, T. J Chem Phys 2002, 117, 8573.
29. Huang, J.; Zhang, Q.; Schlick, T. Biophys J 2003, 85, 804.
30. Silverman, M.; Simon, M. Cell 1980, 19, 845.
31. Johnson, R.; Simon, M. Cell 1985, 41, 781.
32. Heichman, K.; Johnson, R. Science 1990, 249, 511.
33. Gabdoulline, R. R.; Wade, R. C. J Phys Chem 1996, 100, 3868.
34. Schlick, T.; Fogelson, A. ACM Trans Math Softw 1992, 18, 46.
35. Schlick, T.; Fogelson, A. ACM Trans Math Softw 1992, 18, 71.
36. Xie, D.; Schlick, T. ACM Trans Math Softw 1999, 25, 108.
37. Gilson, M.; Sharp, K. A.; Honig, B. J Comput Chem 1988, 9, 327.
38. Sharp, K.; Honig, B. J Phys Chem 1990, 94, 7684.
39. Sharp, K.; Honig, B. Ann Rev Biophys Chem 1990, 19, 301.
40. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
41. Connolly, M. L. Science 1983, 221, 709.
42. Hasted, J. B. Aqueous Dielectrics; Chapman and Hall: London, 1973.
43. Nicholls, A.; Sharp, K. A.; Honig, B. Proteins 1991, 11, 281.
44. Zhang, Q.; Schlick, T. VIEWMODEL 2001. (http://monod.biomath.nyu.edu/~qzhang/Softwares.htm)