

Treatment of counterions in computer simulations of DNA

Ganesan Ravishanker, Pascal Auffinger, David Langley, Bhyravabhotla Jayaram, Matthew Young, David Beveridge

▶ **To cite this version:**

Ganesan Ravishanker, Pascal Auffinger, David Langley, Bhyravabhotla Jayaram, Matthew Young, et al.. Treatment of counterions in computer simulations of DNA. Reviews in Computational Chemistry, John Wiley & Sons, Inc., pp.317-372, 2007, Reviews in Computational Chemistry, 10.1002/9780470125885.ch6 . hal-03420750

HAL Id: hal-03420750
https://hal.science/hal-03420750

Submitted on 3 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CHAPTER 6

Treatment of Counterions in Computer Simulations of DNA

Ganesan Ravishanker,* Pascal Auffinger,†
David R. Langley,‡ Bhyravabhotla Jayaram,§
Matthew A. Young,* and David L. Beveridge*

**Department of Chemistry, Wesleyan University, Middletown, Connecticut, †IBMC, 15 rue René Descartes, 67084 Strasbourg, France, ‡Pharmaceutical Research Institute, Bristol-Myers Squibb Company, 5 Research Parkway, P.O. Box 5100, Wallingford, Connecticut 06492-7660, and §Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India*

INTRODUCTION

Reliable atomic level descriptions of the structure and dynamics of biological macromolecules are important in unraveling the mysteries of biological processes. The recent surge in structural characterization[1,2] of biological systems, such as proteins, DNA, and protein–DNA complexes, has contributed to a better understanding of these processes. Experimental techniques such as X-ray crystallography and NMR spectroscopy are routinely used to study the structure of biological molecules. Computer simulation studies are fast becom-

Reviews in Computational Chemistry, Volume 11
Kenny B. Lipkowitz and Donald B. Boyd, Editors
Wiley-VCH, John Wiley and Sons, Inc., New York, © 1997

ing a powerful adjunct to these methods and provide detailed knowledge of the structure, dynamics, and energetics of biological molecules.[3,4] These studies are resource intensive, requiring fast computer systems and gigabytes of data storage facilities. Affordable, fast UNIX workstations combined with the availability of centralized supercomputing resources at national supercomputing centers have resulted in a tremendous increase in computer simulations of fairly large biomolecular systems, providing detailed insight into some of the fundamental issues in biochemistry and biophysics.

All computer simulations, however, are fraught with numerous methodological difficulties, such as the accuracy of the force fields used to describe the atomic interactions in the system and the choice of the optimal time step for the molecular dynamics (MD). The most severe approximation used in these calculations involves treatment of long-range electrostatic interactions.[5,6] Biological systems contain mobile counterions to balance the net atomic charges on acidic and basic side chains of proteins as well as on the polyanionic backbone of DNA. Proper treatment of the electrostatic interactions involving these counterions is essential in the development of reliable and acceptable models for these systems. We present a review of the methods used in treating counterions in computer simulations involving DNA.

DNA at physiological pH contains a fully charged polyanionic phosphate backbone,[7] which is neutralized by cations such as Na^+, K^+, Mg^{2+}, and/or Ca^{2+}. In addition, the DNA is surrounded by other ions from the salts, controlling the ionic strength of the medium. All these ions are mobile because they are not covalently bound to the DNA. The collection of mobile ions and the solvent near the surface of the DNA are collectively referred to as the "ion atmosphere." The structural characterization of the DNA and its ion atmosphere is one of the most complex problems in structural biochemistry.

The structure of DNA itself has been the subject of numerous studies.[8,9] Fiber diffraction[10–12] and X-ray crystallographic studies[9] have been used to classify DNA structures into distinct classes based on the helical and morphological properties of the DNA double helix. These are commonly referred to as the canonical forms of DNA, the familiar A, B, and Z forms being well characterized. Several other subclasses have also been formulated.[13] Ionic strength of the medium and the water activity near the surface of the DNA, which are the two major components of the postulated ion atmosphere, have been shown to control the structural transitions among the canonical forms of DNA.[7] Structural transitions from the B to the A form is induced by reducing the water activity, whereas the B-to-Z transition is induced by increasing the ionic strength of the medium.

Several theoretical as well as computer models have been developed to promote understanding of the ion atmosphere around the DNA and its effect on the structure and dynamics of DNA.[14–18] Theoretical models usually employ a coarse-grained description of DNA, as a line of charges or a charged cylinder, in the quest to understand the macroscopic nature of the ion atmosphere. Although such methods are a first step to understanding the electrostatics of the ion atmosphere, they have limited practical use because the structure of DNA rarely conforms to the idealized geometries accessible to studies of these types.

Computer simulations, on the other hand, use a fine-grained model in which the ion atmosphere is treated at the atomic level. These methods are desirable because they produce a detailed understanding of the structure and dynamics of the DNA as well as the surrounding ion atmosphere. However, the long-range electrostatic interactions involving the phosphate backbone with these ions, combined with convergence problems associated with ion mobility, pose a major hurdle in developing successful models for DNA in its ion atmosphere. Systems that have been modeled to date contain a few thousand molecules in a central simulation box measuring 30–40 Å on a side, the box being various shapes, with the potential energies of interaction typically truncated at interatomic distances of 10–15 Å. The electrostatic interaction energy between fully charged ions can be as large as 20 kcal/mol at these cutoff distances, however, severely limiting the usefulness of the systems. Several approaches meant to overcome such methodological deficiencies have been proposed, and this chapter focuses on these models.

Validation is one of the most important requirements of a computer model. Any proposed ion atmosphere model should provide an acceptable structural and dynamical view of DNA. An acceptable qualitative model for DNA is generally considered to be one in which the overall double-helical properties, such as propeller twist and rise, and morphological features, such as groove dimensions and helix axis bending, are maintained along with the Watson–Crick hydrogen bonds. Quantitative validation involves detailed comparison of conformational, helicoidal, and morphological indices with those from the canonical forms and from X-ray crystal structures (when available). Structures of aqueous solutions of DNA derived from theory can also be validated against structural data from two- and three-dimensional NMR experiments. This can be done by calculating the NMR properties from the computer model for comparison with experiment. It is far more common, however, for scientists to derive a model for the entire DNA by means of partial structural information from NMR experiments and computer modeling than to validate completely computer-generated structures.[19–22] The DNA structure is only one aspect of model validation. Consideration of dynamics and energetics of the ions around the DNA as well as the ion atmosphere are all tedious, yet important aspects of validation. We now examine various methods available for characterizing and validating the ion atmosphere model for DNA.

The treatment of counterions around DNA in computer simulations requires a thorough knowledge of several aspects of this problem, including the structure of DNA, computer simulation methodology,[23] electrostatic interactions,[5,6] and analysis techniques.[24] Detailed discussions of all these topics are beyond the scope of this chapter. However, we provide brief overviews of these

methods as needed. In the background section, we review the structure of DNA and Manning's counterion condensation theory. In the methodology section, which follows, we describe methods used to model the effects of counterions and give practical recommendations for the novice. The computer simulation section provides a review of earlier research publications on DNA systems in chronological order. The concluding section of this chapter predicts the future of DNA simulations involving counterions.

BACKGROUND

Structure of DNA

DNA's structure has been the subject of many studies since the landmark discovery of the double helix by Watson and Crick.[25] X-ray crystallography was the only technique available to derive atomic models for large biomolecules until recently. The difficulties associated with generating the high quality, single crystals required for X-ray diffraction slowed progress. Arnott and co-workers [10–12] used fiber diffraction methods to evaluate the DNA helix morphology. X-ray diffraction patterns of fibers drawn from highly concentrated solutions of nucleic acids were used to derive useful helix parameters such as the pitch height and number of nucleotides per turn. At least two major helical forms of DNA, the canonical A and B forms, were characterized by this method. Subsequently, several other forms of DNA were catalogued by this procedure.[26] X-ray structures of the monomeric nucleotide units[27] in combination with the helix parameters from fiber diffraction can be used to generate the canonical forms of DNA of any desired sequence and length. Canonical DNA generated this way provides the initial structure for DNA in many computer simulations.

Wang et al.[28] reported the first X-ray crystal structure of a left-handed Z form of DNA. Dickerson and Drew[29] subsequently reported the structure of a dodecamer sequence d(CGCGAATTCGCG) in 1981. This sequence contains the target for the enzyme EcoRI endonuclease and crystallizes in the B form. Interesting structural features, such as helical axis bending, groove narrowing, and sequence-dependent helicoidal parameter variations, emerged from this crystal structure, which is now known as the Drew–Dickerson crystal structure. Other crystal structures of oligonucleotides in their free state, as well as bound to drugs or proteins, have also been solved. A central repository for nucleic acid structures, the Nucleic Acid Database (NDB),[1] was created in the early 1990s and now contains more than 190 DNA and 30 RNA crystal structures.

Complexities related to structural definitions of DNA became evident with the emergence of atomic level information. Whereas the backbone conformation of DNA has been standardized in the form of a complete set of backbone dihedral angles in IUPAC notations,[7] the helicoidal description is so complex that there are at least three different prescriptions for its analysis.[30] The mathematical description of a consistent helical axis system, needed for measuring various helicoidal parameters, is complicated by local variations in the base pair geometry. Generally, the helicoidal analysis of DNA is classified into local[31,32] or global axis systems.[33] Helix morphology is another important description of the DNA structure. Both the quantitative measure of groove structures[24,34–36] and helical axis bending[37,38] fall into this category, and again, this is a complex problem in a molecule as flexible as DNA. Attempts are being made to define these parameters in a consistent way, to be useful for structural comparisons.

NMR spectra of DNA are extremely useful for deriving the solution phase structures of DNA.[19–22] Two-dimensional spectra of DNA obtained by means of nuclear Overhauser effect spectroscopy (NOESY) contain structural information related to interproton distances within 5 Å. Characteristic spectral patterns for known canonical forms of DNA can be used to characterize unknown DNA material. Quantitative atomic models for DNA cannot be derived from NOESY spectra alone because those spectra contain only partial structural information. Additional NMR data, including the sugar proton coupling constants, are used to infer the furan ring puckering, and, in some rare cases, ^{31}P spectra have been used to infer backbone geometries. A powerful technique called restrained molecular dynamics,[22,39,40] in which NMR data serve as restraints in the performance of MD simulations, has been used to derive atomic level description of solution structure of DNA.[41] This technique is both force field and MD protocol dependent, but it provides an ensemble of structures satisfying the imposed NMR restraints. Nonetheless, the deposition of such NMR-derived structures into databases, such as NDB, is becoming common.

Counterions in DNA: Counterion Condensation and Manning Theory

Some of the most successful computer simulation studies of nucleic acid systems to date employ models without explicit treatment of counterions, using instead the Manning model for counterion condensation.[14] Manning developed the counterion condensation model from polyelectrolyte theories applied to DNA treated as a charged cylinder. This model, resulting from a series of studies[14,42,43] is summarized by:

$$\xi = \frac{q^2}{\epsilon kTb} \qquad [1]$$

where the dimensionless parameter ξ is a function of the charge on the counterion q, temperature T, solvent dielectric constant ϵ, and the average spacing

between backbone phosphates along the cylindrical axis, b; k is the Boltzmann constant. The magnitude of net charge on each phosphate due to counterion screening is then given by $1/(N\xi)$, where N is the valency of the counterion. ξ reduces to $7.1/b$ (in Å) for solvent water ($\epsilon = 80$) at 25 °C. For canonical B-DNA, b is taken to be 1.7 Å. The net charge on phosphate groups of a canonical B-DNA is thus −0.24 in the presence of Na^+ counterions and −0.12 for Mg^{2+} counterions.

Manning's theory illustrates that the net charge on a phosphate group due to counterion condensation is independent of the ionic strength of the medium, a surprising yet important result. Though the Manning model does not specify structural characteristics of the condensed counterions at atomic resolution, an important observation about the ion atmosphere was made in subsequent studies and supported by ^{23}Na NMR experiments.[15,44–46] Specifically, "condensed" counterions, defined as those within ~17 Å of the helical axis of the DNA (referred to as the Manning radius), were shown to give rise to the possibility of both contact and solvent-separated ion pairing.[14] These counterions are not necessarily bound to any single phosphate. The noncondensed ions are interpreted as forming a diffuse cloud outside the Manning radius. Manning applied the counterion condensation theory to derive various observed macroscopic properties of DNA.[14,43,47]

Fenley, Manning, and Olson extended the counterion condensation theory with a more realistic representation of DNA treated as a three-dimensional, discrete charge distribution.[48] General agreement with experimental observations is obtained by means of a linear lattice model for DNA in a uniform dielectric or a charge distribution corresponding to the backbone phosphate geometry in B-DNA combined with a distance-dependent dielectric. Dewey developed a counterion condensation model for oligoelectrolytes at high ionic strength.[49] Several recent computer simulation studies have applied the Manning model to the simulation of DNA, whereby the net atomic charges of the phosphate group are adjusted to represent the fraction of the condensed counterion. These studies are discussed in detail below.

Several other theoretical approaches, including solutions to Poisson–Boltzmann equations,[50] quantum mechanics,[51] and integral equation methods,[52–54] have been used to characterize the electrostatics of DNA. Detailed discussion of these methods can be found in the literature[50–54] and references therein. Most of these studies find that the largest electrostatic potentials for DNA are in the grooves, a property that can be useful when validating and characterizing simulation results.

Although the results on counterion distributions in nucleic acid systems from X-ray crystallography are fragmentary,[55] there exist several NMR studies about metals in DNA systems.[15,44–46] The data from ^{23}Na NMR experiments proved to be difficult to interpret. The quadrupolar relaxation mechanism of this nuclide, originating from the electric field gradients, obfuscates the results, but in general the data are consistent with the counterion condensation theory.

Bleam, Anderson, and Record[15] discuss a quadrupolar relaxation mechanism arising from the radial diffusion of ions that is DNA conformation dependent, but sequence independent. However, a recent theoretical study by Reddy, Rossky, and Murthy[17] demonstrates that relaxation is dominated by the motion of ions in the vicinity of the DNA backbone, where the electrostatic potential is the greatest, rather than by radial diffusion. Bacquet and Rossky,[56] used the hypernetted chain (HNC) formalism to calculate the mean-square electric field gradients of ion distributions around DNA. (HNC is an integral equation method based on graph theory and is used to evaluate correlation functions.) Their results were in good accord with those estimated from experimental NMR line widths. This study used a two-state model for the counterions, consistent with earlier proposals in this regard.[15,44] The two-state model assumes that counterions in the system can be divided into two distinct groups, bound and free, with different mean relaxation rates. The bound ions lie within a few hydrated ionic radii of the DNA surface and experience the largest electric field from the polyanionic backbone of DNA, whereas ions not influenced by this field are considered "free."[15]

Bleam, Anderson, and Record[15] used the two-state model for the counterions around DNA to assess the observed invariance of condensed counterion concentration with changes in ionic strength. This invariance, recall, is consonant with the Manning model. It was observed that NMR line widths correlate linearly with ionic strength. This correlation in combination with the assumed two-state model implies the product $r\,(R_B - R_F)$ is constant, where r is the number of bound counterions per phosphate group and $(R_B - R_F)$ is the difference between NMR relaxation rates of bound and free ions. Assuming that these two factors are individually constant, the fraction of condensed counterion can be estimated, ranging from 0.65 to 0.85 in this study but around 0.53 in other work.[46] Recent NMR studies [57–62] using divalent cations like ^{25}Mg, ^{43}Ca, and ^{59}Co show that multiple binding environments in B-DNA for these ions may exist. Experimental studies of salt effects on the helix–coil transitions of oligonucleotides demonstrated that 0.08–0.13 Na^+ ion per phosphate is released upon helix melting,[62,63] and this fraction was sequence dependent.

Computer simulations on DNA without explicit counterions use different implementations of Manning's counterion condensation theory. Most of them simply scale down the net atomic charges on the phosphate group (generally O5′, P, O1P, O2P, and O3′) by a factor of 0.24–0.34. This scaling tends to reduce the phosphate charges to such an extent that some other atoms in the DNA have relatively high charges. The CHARMM force field[64] uses a different approach, whereby the charges of individual atoms in the group are still high, but sum up to −0.34. In at least one study,[65] the charges on the entire DNA were scaled down by 0.25; consequently, the counterion condensation was assumed to affect the entire DNA rather than the backbone phosphate group alone.

METHODOLOGY

Computer Simulations: An Overview

An atomistic computer simulation requires an atomic level description of the system along with the specification of the simulation protocol. Such a description, which usually includes the force field representing all interactions in the system for such large molecules and their environment, is collectively referred to as the "model." A simulation "engine," typically a Monte Carlo or a molecular dynamics algorithm, then acts on this information to probe the system in configurational space. The computer programs output the configurational states of the system generated during the course of the simulation, commonly referred to as the "trajectory" in the case of molecular dynamics simulations. Analysis of these states is used to validate the computer model for the system and subsequently to derive new and useful inferences about the structure and dynamics of system.

Detailed descriptions of common simulation engines for atomistic modeling can be found in several excellent sources.[23,66,67] We briefly describe the two most commonly used techniques because of their relevance to this chapter. The Monte Carlo (MC) method is typically used to explore configurational space where all internal motions of the molecules are "frozen," allowing translations of mobile ions and translations and rotations of explicit solvent molecules to generate Boltzmann-weighted configurations. Though Monte Carlo methods have also been applied to study systems executing all internal motions, methodological difficulties associated with such simulations make it impossible to apply them to fairly large solute molecules as in this chapter.

Molecular dynamics methods can be used to explore the isoenergy phase space of molecular systems. Coupled Newton's equations of motion for each atom in the system are solved to generate the trajectory for each atom moving in the force field of all other atoms:

$$\frac{d^2\mathbf{r}_i(t)}{dt^2} = \frac{1}{m_i}\mathbf{F}_i \qquad [2]$$

$$\mathbf{F}_i = -\frac{\partial V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)}{\partial \mathbf{r}_i} \qquad [3]$$

Here $\mathbf{F}_i$ denotes the force on atom i due to all other atoms and V is the potential energy of the system. Generating configurations in a MC simulation is determined by the Boltzmann weight of the total energy of the system in a stochastic manner, but configurations generated in an MD simulation are controlled by the net force acting on each atom in the system making this a deterministic approach.

MC simulations produce a set of configurations with various energy-weighted arrangements of mobile ions and solvent molecules around the solute.

This information can then be analyzed by means of pair correlation functions to determine structural indices. In contrast, MD simulations provide a *trajectory*—that is, an evolution history in time—with the atomic motions on the femtosecond time scale. One or more of these trajectories can be examined to derive motional characteristics (dynamics) of the solute, as well as to assess interdomain communications within the solute. Thermodynamic properties can be derived from both MC and MD simulations by calculating the ensemble average and time average, respectively, of these quantities. The time average calculated from MD and the ensemble average from MC should converge according to the ergodic hypothesis.[66] MD simulations can be used to estimate dynamical properties of the system not generally accessible in MC simulations. Complete overviews of these methods can be found in traditional textbooks.[66]

System Description

A system description involves defining the composition and spatial arrangement of all particles, including the number of solute molecules, mobile ions, and solvent molecules. Although simulation of systems containing mixed solvents and several solute molecules is conceivable, the task is made difficult by the size of such systems, combined with the methodological difficulties associated with mixed solvent simulations.

The spatial arrangement of solute heavy atoms of biological molecules is available from a wide variety of sources, especially those from X-ray crystallography. The Brookhaven Protein Data Bank (PDB)[2] contains over 600 sets of coordinates for proteins, and NDB[1] contains over 200 crystal structures of oligonucleotides. X-ray crystal structures of protein–DNA complexes are also available from both sources. Increasingly, solution structures of DNA and protein, solved by NMR spectroscopy, are becoming available. Structures of canonical DNA can also be derived by means of fiber diffraction data.[11]

An initial structure of the DNA requires hydrogen "capping" at the 5′ and 3′ ends. Figure 1 shows a two-dimensional representation of a sample nucleotide along with the atom names. The 3′ cap is unambiguous in that a hydrogen, H3′, is added to the O3′ atom. However, the 5′ cap has traditionally been handled two different ways, one with and one without the terminal phosphate group. In the former case, the cap is an H3′ atom on the terminal phosphate O3′, whereas in the latter, it is an H5′ on the O5′. The atoms at the capped ends have their charges adjusted to be different from those of the same kind in the interior of the DNA. This is because the chemical environments of the termini and interior differ. When the phosphate cap is used, the number of phosphate groups per strand must match the number of residues; with the O5′ cap, however, it will be one less. Both types of cap have been used in DNA simulations. Because electrostatic interactions are so hard to model, one may take the view that "less is better" and choose the O5′ cap strategy.

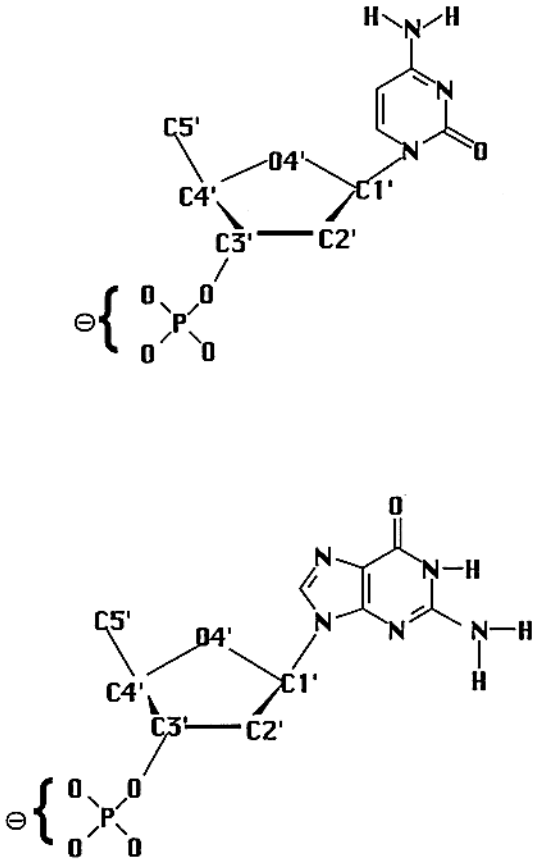


Figure 1 A two-dimensional representation of a nucleotide fragment with the associated atom labels.

In simulations involving explicit ions and solvent molecules, an initial *configuration* (conformation of the solute plus the positions of the associated solvent molecules and ions) is selected by guessing because the necessary experimental data do not exist. This is an important step when simulating systems containing counterions, especially with MD techniques, but presumably not for MC. Computer simulations should fully explore the vast unbounded configuration space of the system. Unfortunately, the simulations are of finite length, and the configurational space spanned during the course of the simulation is highly dependent on initial conditions.[68] The MD trajectories resulting from different initial configurations are more likely to diverge than converge. Therefore extreme caution is required when the initial configuration of an

aqueous solution model is specified for MD. A workaround is to run several short simulations using different initial configurations and treat the configurations from all these trajectories when constructing the ensemble average of properties.[68]

Placement of Ions

Several techniques have been used for placing counterions around the DNA. The simplest involves first placing solvents around the DNA and then computing the electrostatic potential of the DNA on each solvent water. The water molecules experiencing the highest negative potential are then replaced by Na^+ or other metal ions. How this is done, of course, depends on the size of the counterion relative to the solvent water it replaces; for larger ions, two or three waters need to be replaced, sometimes creating a vacuum in the system. Another method involves calculating the electrostatic potential around the free DNA embedded in a predefined grid followed by the placement of counterions at the grid points with the largest negative potentials. This has the effect of generating an energetically unfavorable initial configuration because two counterions can be too close to each other, especially in irregular DNA or highly bent DNA. The problem can be overcome by placing the first ion at the grid point with the deepest potential, and then recalculating the potential with the already placed ions and iteratively placing the remaining ions.[69] This seems to be the optimum strategy for placing ions using the electrostatic potential. An initial configuration of 22 Na^+ ions placed around the dodecamer d(CGCGAATTCGCG) using this method is presented in Figure 2. Each ion is tagged with its order of placement. These methods have the advantage of being driven by electrostatics; moreover, the geometry of the DNA influences the placement.

Other methods for placing explicit counterions are based entirely on the phosphate group geometry. The most commonly used method places the ions along the bisector of the O1P—P—O2P angle at a distance 4.5–6.0 Å from the central P. These distances reflect the contact and solvent-separated ions, respectively. An initial configuration of 22 Na^+ ions around the canonical B-form Drew–Dickerson dodecamer at 6 Å from the phosphate bisector is shown in Figure 3. Placement of ions along the P–O1P or P—O2P vectors also has been tried. MC simulation of these initial systems to equilibrate the positions of counterions almost always leads to energetically stable configurations in which the counterions move into positions in between two phosphate groups. This result suggests an alternate placement method of placing counterions along the P—O5′ (or P—O3′) vector at a distance of 6.0 Å. This type of placement is further justified by the fact that in certain protein–DNA complexes, the charged side chains of lysine and argenine are found in such positions.[1]

Whatever method is implemented, there must be sufficient equilibration to allow the ions to occupy optimal positions around the DNA. Typical strate-

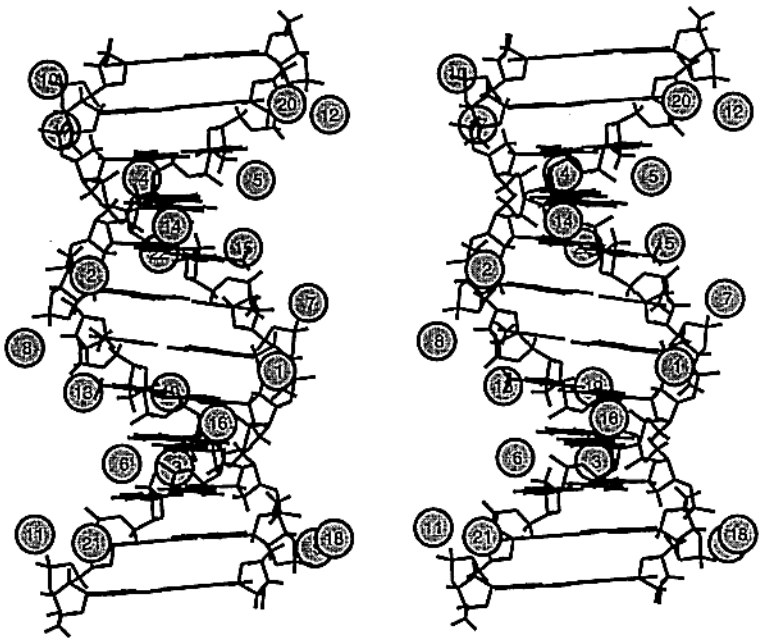


Figure 2 Stereo diagram showing the positions of Na^+ around d(CGCGAATTCGCG) in its canonical B form. The counterions were placed using the "cion" module from AMBER 4.1 (Ref. 69). This program calculates the potential due to the DNA and other ions in predefined grids around the DNA and places the counterions in grids of greatest potential. The order of placement of the ions is indicated.

gies for equilibration involve performing extensive MC on ions and water prior to an MD calculation, or keeping the DNA fixed with positional restraints while performing MD on the ions and water.

Placement of Water Molecules

Specifying the arrangement of water molecules around a solute is done in several ways. Liquid water simulations[70,71] carried out to verify potential functions were begun from boxes consisting of six layers of water molecules in all three dimensions, giving rise to a box containing 216 water molecules. Most of the computer simulation programs[64,69,72] provide a preequilibrated "box" of water molecules, typically a cube, with 216 water molecules. This produces a cubic box roughly 18.5 Å on a side, allowing a 9 Å cutoff for water–water interaction, which can be used in testing the potential function. Images of the 216-water box are propagated in three dimensions, generating a larger volume of desired size and shape. The large box created in this way serves as the central

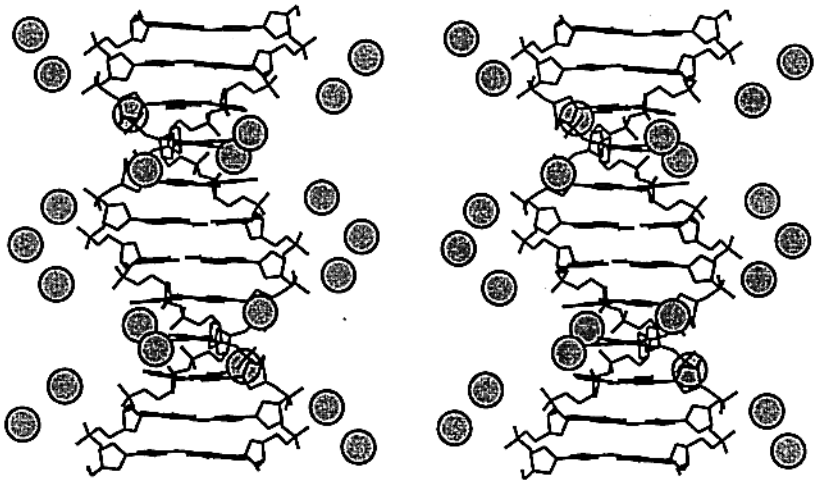


Figure 3 Stereo diagram showing the placement of counterions 6 Å from the phosphorus along the O1P–P–O2P bisector. The same canonical B-form structure of DNA used in Figure 1 was used here.

box for the simulation. The solute is then placed at the center of the central box, and all water molecules within the van der Waals radius of any solute atom are removed. Ions are then placed at positions as described in the preceding section, and any waters within the van der Waals radius are also removed. This procedure is likely to give a solvent density less than that of a pure water box, which is 1. However, the solvent bath being constructed is so large that the number of water molecules removed for reasons of van der Waals overlap is small in comparison, and the resulting water density is usually close to 1. The water structure in the initial configuration is biased toward the bulk water, and long equilibration runs are required for the waters to readjust to the influence of DNA and ions.

Another approach[73] is to first place the solute molecule and ions in the simulation box. The box is then divided into cubes comparable in size to the diameter of the water molecule, and water molecules, in random orientation, are placed into each cube not containing portions of solute or ion. This approach creates larger "voids" near the surface of the solute molecule because cube division always begins at one face of the box. An attractive alternative for regular-shaped solutes (e.g., canonical A-, B-, Z-DNA) is to extend the cubes outward from the surface of the solute. This alternative requires more equilibration than is needed to start from a preequilibrated 216-water box because of the inherent order in the system created by ordered cubes. Unlike the method featuring a preequilibrated water box, the latter approaches need time to establish intermolecular water structure.

Force Fields

Monte Carlo and molecular dynamics sampling methodologies usually require the energetics of all pairwise interactions in the system. In addition, MD requires the energetics of internal coordinate motions and calculation of forces, which are the analytical first derivatives of the potential energy. The list of internal coordinates required for MD simulations typically is constructed by means of predefined monomeric residue topologies, usually provided as a part of the modeling program. The complete description of all energetic terms in a system is referred to as the force field for the system. Force fields for biological molecules developed over the past 15 years include CHARMM,[64] AMBER,[74] OPLS,[75,76] and GROMOS.[72] Constant refinement of the earlier versions of these force fields has resulted in a more robust and reliable set of parameters for use in computer simulations. Detailed discussion of force field is beyond the scope of this chapter, and the reader is referred to the original papers describing the force field development.[64,72,74–76]

Energy

A system's nonbonded molecular mechanics energy is typically calculated by means of the pairwise additivity assumption. (The system's total molecular mechanics energy would also include the valence terms, such as bond stretching, bending, and torsion.) Only interaction energies between all applicable pairs of atoms are computed, three-body, four-body, and higher order interaction terms are ignored. Effective pair potentials can be employed this way because judicious parameterization incorporates the cooperative higher order effects to some extent. The total potential energy expression consists of terms arising from internal coordinate motions a well as nonbonded interactions. The long-range nonbonded interactions accounting for dispersion and electrostatic interactions usually contribute over 90% of the total energy, especially in charged systems. The electrostatic interactions fall off as $1/r$ and the dispersion terms arising from dipolar interactions fall off as $1/r^3$. Evaluating these terms is the most time-consuming aspect of computer simulations, accounting for 90–95% of the computer time.

The relatively short range of dipole–dipole interactions, compared to electrostatics, provides the rationale for evaluating electrostatic energies using a concept called "neutral charge groups."[72] Interactions between neutral charge groups can be approximated as being r^{-3} dependent.[23] The neutral charge group scheme identifies functional groups in the solute that are charge neutral, (i.e., the sum of net atomic charges of all atoms in a given functional group is zero). If two neutral charge groups qualify as an interacting pair within a cutoff criterion applied to the geometric centers of the groups, pairwise interactions between all atoms in the two groups are included explicitly, even though some of these atoms may be beyond the cutoff distance. In some cases,[69] an entire

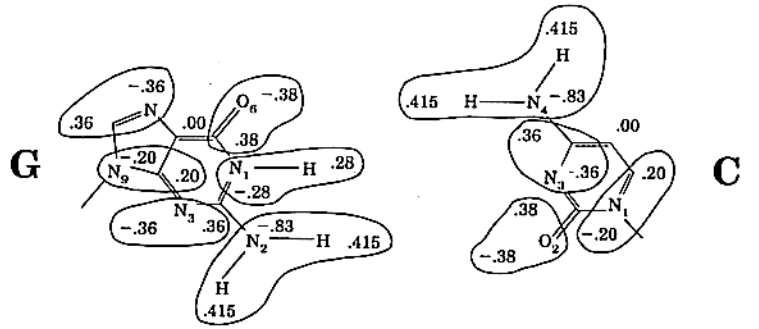


Figure 4 Neutral charge groups in a nucleotide from GROMOS. Groups are indicated by a line enclosing the atoms in each group.

residue is treated as a single charge group. Here, however, the decision to calculate the interactions is not based on the geometric center of the residue, but rather on the minimum distance between any pair of atoms between the two residues. That is, as long as there exists even a single atom pair within the cutoff distance, all interresidue pairwise interactions are included. It is obvious that the energetics based on atom pairs and the energetics based on charge groups produce different numbers of interactions and therefore are not equivalent. The choice of atoms used to form a neutral charge group plays a key role in the description of the system, and difficulties associated with the application of this concept to noncovalently bound atoms, especially mobile counterions, are discussed at great length in the following sections. Figure 4 shows two examples of charge groups for DNA fragments defined in GROMOS.

SIMULATION PROTOCOLS

Ensembles

Simulation protocols begin by specifying the statistical ensemble used for the system. The four most commonly used statistical ensembles in this area of research are the microcanonical (E, V, N), the canonical (T, V, N), the isothermal–isobaric (T, P, N), and the grand canonical (T, V, μ) ensembles, where E is total energy, N is the total number of molecules, T is the temperature, P the pressure, V the volume, and μ the chemical potential of the system. Depending on the chosen ensemble, the three listed thermodynamic quantities are required to be conserved throughout the simulation. All these ensembles are implemented with both MC and MD methods.

The grand canonical ensemble, with its algorithmic complexity and convergence problems, is the least used. The canonical ensemble is the easiest to

implement. The microcanonical ensemble is more appropriate for MD simulations where, by definition, the total energy of the system has to be conserved. The canonical and isothermal–isobaric ensembles are used in both MC and MD simulations. The choice of an ensemble is driven by the goals of the researcher and the nature of the computer resources available. Statistical mechanical equations exist for the interconversion of the calculated thermodynamic quantities from one ensemble to the other. Allen and Tildesley[66] provide a thorough discussion on the various ensembles, their implementations, and advantages and disadvantages.

System Environment

The next step is to consider whether the simulation should be set up in vacuo, in a crystalline environment, in a "cluster," or in solution. In the first case, ions may be included, but constraints must be applied to keep them from "evaporating" (i.e., drifting too far from the solute). The in vacuo method has limited use. Clusters are generally solutes and ions surrounded by a sheath of solvent molecules large enough to provide a reasonable solution environment. There is usually a need for at least three solvent layers around the solute, and preferably an additional shell for large biomolecular systems to account for any unforeseen deformations in solute conformation. Unbalanced forces at the surface of the cluster may cause the solvents to evaporate, so additional restraining forces may be required to keep them in the cluster.

Solution environment is established by the application of periodic boundary conditions (PBC), whereby the replication of the central simulation box in all directions provides, in essence, an infinite system. This approach is most appropriate for studying the solution structure of biological molecules, especially those involving mobile counterions. The computational complexity increases severalfold when going from in vacuo to cluster to solution simulations, because the number of pairwise interactions increases tremendously.

Periodic Boxes and Cutoff Distances

The central stimulation box can exist in any geometrical form, as long as the application of periodicity to generate images of this box is computationally viable. Because a significant chunk of computer time is spent on evaluating pairwise interactions between the atoms in adjacent images, care must be taken in choosing and implementing the periodic boundary conditions. Most simulation programs provide simple cubic (SC) and face-centered cubic (FCC) periodic boundary conditions, both of which require a central box that is a cube. Rectangular boxes are a natural extension to SC. Although these boundary conditions are easy to implement, they may not be the most appropriate for nonspherical solutes like DNA. Generation of a cubic box of water for such systems adds many more waters than necessary along the short principal axis

directions of the molecule. In the interest of economy, hexagonal prism (HP) boundary conditions have been implemented. For a given density, the number of water molecules required to provide a minimum 9 Å solvent layer around every atom in the molecule requires far fewer waters in an HP box than in a SC box. The HP box provides tremendous savings in computer time.

Under periodic boundary conditions, when an atom in the system leaves the central box, an image of it from one of the adjacent boxes must enter the system from the opposite face to conserve N. An adjunct to periodic boundary conditions is the minimum image convention, whereby any atom in the central box is assumed to interact only with the closest periodic images of all other atoms in the central box. The minimum image convention requires a total of $N(N-1)/2$ pairwise interaction calculations at every MD step, or $(N-1)$ for each MC step. Systems consisting of several thousand atoms can be very costly to implement in this way, and spherical cutoffs are applied to reduce the number of interactions by truncating the nonbonded interaction energy calculation for each atom at large enough distances. A spherical cutoff introduces severe problems for ionic systems, however, because individual ions will not see equivalent numbers of opposite charges within the cutoff sphere, and the transition of ions across the cutoff boundary will further worsen the underlying electroneutrality of the system. The implication of this computational artifact is discussed in detail below. However, methods like Ewald summation and use of a reaction field provide alternatives that are appropriate for these systems.

The number of pairwise interactions for any atom is proportional to r^3, where r is the spherical cutoff radius. Historically, r values ranged from 7.5 to 8.5 Å, but the availability of enhanced computer resources has made it possible to study systems with r in the 10–15 Å range. In fact, it has been shown that cutoff distances should be at least as large as 12–15 Å to obtain good results using pairwise potentials.[77] The upper limit for spherical cutoff under minimum image convention is half the box length or, in the case of noncubic boxes, half the length of the lowest dimension. Computing all interacting pairs at each step of the simulation is time-consuming. Therefore, a data structure called the nonbonded pair list, containing all interacting neighbor atoms within the cutoff range, is updated every few steps of the simulation. Details regarding the nonbonded pair list, such as the optimal data structures and algorithms for updating the lists, can be found elsewhere.[23,66]

Alternative truncation schemes use different cutoffs for different parts of the energy evaluation, thereby introducing heterogeneity into the system. For example, all solute–solute interactions may be calculated using a cutoff different from that for solute–solvent interactions or for solvent–solvent interactions. Smooth truncation schemes provide an attractive alternative to the abrupt truncation using spherical cutoff. Switching and shifting functions[72,77] fall into this class and are discussed in detail below. Twin-range cutoffs employ two types of truncation, an interior and exterior cutoff, whereby the interactions of all atoms in the interior cutoff range are calculated at every step of the

simulation, while all others in the region between the interior and exterior ranges are recomputed only when the nonbonded list is updated. This is equivalent to having a constant field arising from the layer of atoms between the two cutoff ranges between the update times. Characterization of the truncation schemes and their effects on the calculated properties can be found in several excellent sources.[5,6,77,78]

Application of PBCs to systems containing neutral charge groups is cumbersome. It is conceivable that charge groups need to be split at the box edges, with some atoms inside the central box and some outside. Water atoms are considered to be part of a neutral charge group, yet it is very common to find split waters at the edges of the central box. Elimination of this problem requires reconstruction of the entire charge group prior to energy evaluations. Special computer codes needed to handle this are implemented in several simulation programs.[64,69,72] In general, the cutoff criterion is applied to the geometrical centers of neutral charge groups.

It is appropriate at this point to discuss implementation of the PBCs and their customization to emulate different systems using the same central box. Retaining the coordinates of all the molecules in all the image boxes is not desirable because it requires too much computer memory. Moreover, updating the values of all these coordinates at every simulation step takes too much CPU time. Generally, then, the molecules and ions in the central box are allowed to move freely during the course of the simulation, and the identity of the central box is blurred. After some number of simulation steps, the coordinates of the molecules will be spread over the image boxes. For convenience, some programs recreate the central box after several simulation steps. Under these circumstances, an elegant way to handle the PBC for a simple cubic box is to simply correct each of the x, y, and z coordinates of atoms outside the central box by subtracting the appropriate box lengths. More complicated correction schemes are required for HP boundary conditions.

In principle, one can choose to apply the PBC and minimum image correction only to selected distances in the system. The energy evaluation is then usually partitioned into solute–solute, solute–solvent, and solvent–solvent contributions, with the mobile ions almost always considered to be part of the solute. If the PBC is applied to all these interactions, the system being studied is an exact replication of the central box. The concentrations of DNA and counterion can be evaluated from the number of molecules in the box and the volume of the box. On the other hand, one can choose not to apply the minimum image criteria to solute–solute interactions. The result is a system that is effectively less concentrated than the PBC model and may be desirable for DNA solutions. However, special care must be exercised when treating the mobile ions. In the worst-case scenario, if ions are allowed to move freely, two of them may simultaneously reach opposite faces of the central box at the contact distance. If this happens, the ion–ion interactions will be confined to the central box, not the minimum image distance, and because they are outside

the solute–solute cutoff range, their interaction will not be calculated. An elegant approach to this difficulty is to treat the ions as part of the solvent. Unfortunately such a solution becomes a programming nightmare. If this approach is chosen, the ions must be restrained, to ensure that they remain near the surface of the DNA.

Time Scale

Specification of all the above-mentioned components of the computer simulation precedes use of the simulation engine. Sometimes additional declarations need to be provided (stochastic boundaries, constraints, restraints, etc.). MC simulations usually are set up with the solute fixed in the center of the box, allowing only random moves of a solvent molecule or ion. Methods such as force-biasing or preferential sampling effect improvements in sampling efficiency by means of special techniques to modify the direction of motion or the choice of which solvent or ion to move. Each new configuration, after the move of a solvent or ion, is accepted with the Boltzmann probability. The translation and rotations are customized so that 40–60% of the configurations are energetically acceptable. The resulting set of configurations and their energies are then used to calculate statistical thermodynamic quantities as well as to derive conclusions regarding the overall deployment of ions and water around DNA. In MC simulations, then, we need not worry about a time scale. In MD simulations, however, new positions and velocities are generated for each atom after each time step by means of numerical integration of a set of simultaneous differential equations of motion. Time steps in the range of 1–2 fs are most common, with simulation time lengths now in the hundreds of picoseconds to nanoseconds.

There are generally two parts of an MC simulation: equilibration and production. Equilibration is the initial phase of the simulation, where the high energy initial configurations are allowed to settle into local minima on the potential energy surface. The production phase is the postequilibration part, where the ensemble averages of the system are collected.

In MD, there are generally four steps in the simulation: minimization, heating, equilibration, and production. The first phase involves a simple energy minimization (molecular mechanics) to bring the system to a local minimum. Velocities are then assigned to each atom selected from a Maxwell distribution to slowly increase the system's kinetic energy until the target temperature is achieved. Equilibration follows, in which velocities corresponding to the target temperature are reassigned constantly, and finally one enters a production stage of the simulation. There are no predefined recipes telling how long a simulation should be run. According to the ergodic hypothesis, however, the time average of properties calculated from MD simulations and the ensemble average from MC simulations should converge. This statement may be used as a criterion to determine the length of simulation by comparing properties

calculated from both MC and MD simulations on the exact same system using the same statistical mechanical ensemble. Because such tests are likely to produce system- and protocol-dependent results, the time scale of a simulation is determined solely by the scientific goals of the project and the availability of computer resources. It is common to perform MC calculations involving several million steps, and nanosecond MD simulations of biological systems are becoming common, but the reliability of the force fields at such long time scales has not been determined.

Nonbonded Interactions

In this section, we present a detailed discussion of methods relevant to counterion simulations of DNA. Electrostatic interactions dominate all energies involving the DNA and ions. The largest systems simulated to date typically require truncation of nonbonded interaction potentials at around 15 Å. Abrupt truncation of the potential has the adverse effect of creating an artificial boundary. In an MD simulation, the abrupt truncation produces a discontinuity in force because the first derivative of the interaction potential at the cutoff radius is infinity. Computer programs simply set the forces at the boundary to be zero instead of large values to represent infinity.

Abrupt truncation of potential energy also results in an increase in the kinetic energy (temperature) of the system. This can be understood in terms of the electrostatic energy profile for two like-charged and two oppositely charged ions shown in Figure 5. Consider first the case of the two like-charged ions. In MD simulations, the motions of each atom in the system are determined by the magnitude and the direction of the total force on that atom. The force points in the direction of the local minimum on the potential energy surface; the steepness of the energy well, combined with the current location relative to the well, determines the strength of the force. If we consider two like-charged ions in close contact, they are at the left side of the energy profile (Figure 5). The dynamics of this ion pair can be described in terms of the second ion rolling along the energy profile away from the first, constantly reducing the energy of interaction. This reduction in potential energy is accompanied by an increase in the kinetic energy of the rolling ion, increasing the overall temperature of the system.

Extending this argument to the truncated potential, we see that when the ions separate to the cutoff distance, there is a sudden drop in the repulsive potential energy, resulting in the abrupt acquisition of kinetic energy by the second ion, which in turn leads to an artificial increase in the temperature. It is also clear that once separation beyond the cutoff has occurred, there may not be enough accessible thermal energy to climb back up the hill. The result is a quasi-ergodic condition, whereby the two like ions, once separated beyond the cutoff, will never explore regions of configuration space involving distances less than the cutoff.

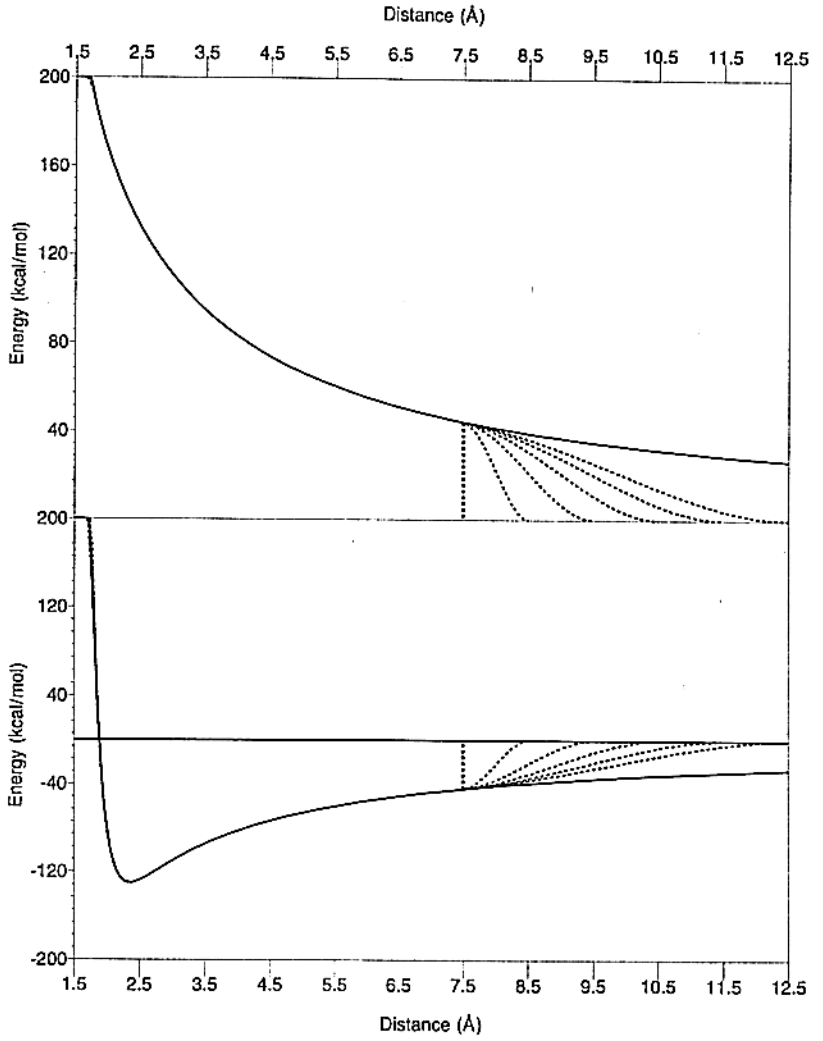


Figure 5 Nonbonded interaction energy profile for two like-charged ions of magnitude +1 (*top*) and two oppositely charged ions of magnitude +1 and −1 (*bottom*). A van der Waals term commensurate with Na^+ and Cl^- is included in the energy. The thick vertical lines represent the energy function with abrupt truncation at 7.5 Å. The dashed lines represent switching functions with r_{on} at 7.5 Å and r_{off} at 8.5, 9.5, 10.5, 11.5, and 12.5 Å respectively.

The same type of argument can be applied to oppositely charged ions. Here the effect is in the reverse direction. When two such ions are farther apart than the cutoff, the sudden dip in the potential energy profile at the cutoff is favorable and the second ion will quickly roll down the dip, increasing the temperature again. It will continue to roll until contact with the other ion is established. Here, the truncation prevents the oppositely charged ions from ever separating beyond the cutoff distance.

The examples presented in Figure 5 involve simple systems. In simulated systems of aqueous solutions, numerous other factors (e.g., the presence of explicit water) influence the electrostatic interactions. Also, the examples given here assume that the box is chosen to be slightly larger than the cutoff used, an assumption common to most MD practitioners.

Auffinger and Beveridge[79] used MD simulations on a solution of NaCl to demonstrate the effects of truncation. The pair correlation function, $g(R)$, for like ions showed a very large peak at the cutoff distance, an effect explained above. Similarly, there is a significant and sudden depletion in the amplitude of $g(R)$ for the Na^+—Cl^- distance, which is due to the preference of these ions to cluster together and stay within the cutoff distance. Such phenomena are natural outcomes of the truncation of the interactions in simulations involving highly charged components, be it a protein with charged side chains[80] or DNA. Although it is difficult to detect the presence of truncation artifacts in such large systems, and then to prove their existence, the NaCl simulation illustrates that these are indeed artifacts that manifest themselves in computer simulations.

Switching Functions

Switching functions[23,77] provide a simple means of removing the truncation discontinuity. These functions require two parameters, referred to as R_{on} and R_{off}, that define the range of distance over which the potential energy of interaction is taken to zero. The mathematical definition of this functional form is

$$SW(r, R_{on}, R_{off}) = \begin{cases} 1 & r < R_{on} \\ \left[\dfrac{(R_{off}^2 - r^2)^2(R_{off}^2 + 2r^2 - 3R_{on}^2)}{(R_{off}^2 - R_{on}^2)^3}\right] & R_{on} < r < R_{off} \\ 0 & r > R_{off} \end{cases} \quad [4]$$

Here, the function $SW(r,R_{on},R_{off})$ is a multiplicative factor for the interaction energy, and its analytical first derivative provides the force. The simple spherical cutoff truncation can be classified as a switching function where R_{on} and R_{off} are identical. A nontruncated potential is one in which both R_{on} and R_{off} are infinity. The range of switching function should be chosen to minimize the manifestation of simulation artifacts. Some aspects of these switching functions can be understood from Figure 5, which shows the electrostatic potential between two like-charged ions with and without switching functions. In general,

the smaller the range, the steeper the function becomes, and steep functions should be avoided. Also the actual values of R_{on} and R_{off} play an important role. For example, the closer R_{on} is to the van der Waals radii, the steeper will become the function, and again, this result is undesirable. There are no general rules but a good compromise is 7.5 and 11.5 Å for R_{on} and R_{off}, respectively.

In systems with neutral charge groups, a single switching function should be applied to all interacting pairs of atoms.[72] Otherwise, if the switching functions are based on the atom pair distances, the group will be split. The switching function for a neutral charge group is usually based on the distance between the geometric centers of the neutral charge groups, and, for this reason, the atoms constituting the charge groups should be selected carefully.

The neutral charge group model is strictly valid only when the groups are electrically neutral. This poses a major problem for the backbone of DNA. The backbone phosphate group consisting of O5′, P, O1P, O2P, and O3′ is a functional group with net charge of −1.0, hardly a neutral charge group. When using the neutral charge group model, then, how does one handle these phosphates? A simple solution is to treat them with their associated mobile counterion as a neutral charge group. The key to this approximation is that each of the counterions is associated with a single unique phosphate group, requiring their close proximity at all times. This can be achieved only by applying restraints to the P–Na^+ distance. Unfortunately, restrained ions cannot explore phase space effectively under MD conditions.

However, a more severe problem arises from the size of this noncovalent neutral charge group. As depicted in Figure 6, we assume that the R_{on} is 7.5 Å and R_{off} is 11.5 Å, while a harmonic restraint on the P–Na^+ distance prevents the Na^+ from moving more than 6 Å from P. In the worst case, the Na^+ is at 6 Å and the geometric center of the charge group is ~4 Å from Na^+ toward the P. The switching region for this group is represented by the shaded shell. Interactions between Na^+ and other groups, including water molecules, within 7.5 − 4.0 = 3.5 Å, are going to be switched on, and interactions between Na^+ and groups at distances exceeding 7.5 Å (11.5 − 4.0) are completely omitted. This is too unrealistic to be useful. If the switching function range is smaller, the effects become even worse. Therefore, different switching function cutoffs specific to these groups are required, and typically values larger than the usual solute–solute R_{on} values are used.

In DNA simulations, the range of the switching function for solute–solute interactions should be examined carefully. Unusual intra- and interstrand P–P distances can indicate artifacts arising from the switching function. In Figure 7 we provide the P–P distribution for 61 B-DNA crystal structures[1] of various sequences and lengths as well as for a canonical B-DNA[11]. Figure 8 shows the same distribution along with P–Na^+ and Na^+–Na^+ distances from a 500 ps simulation[81] on the canonical B form of the Drew-Dickerson dodecamer using the GROMOS force field.[72] The switching region in this simulation was 7.5–8.5 Å. Explicit counterions were

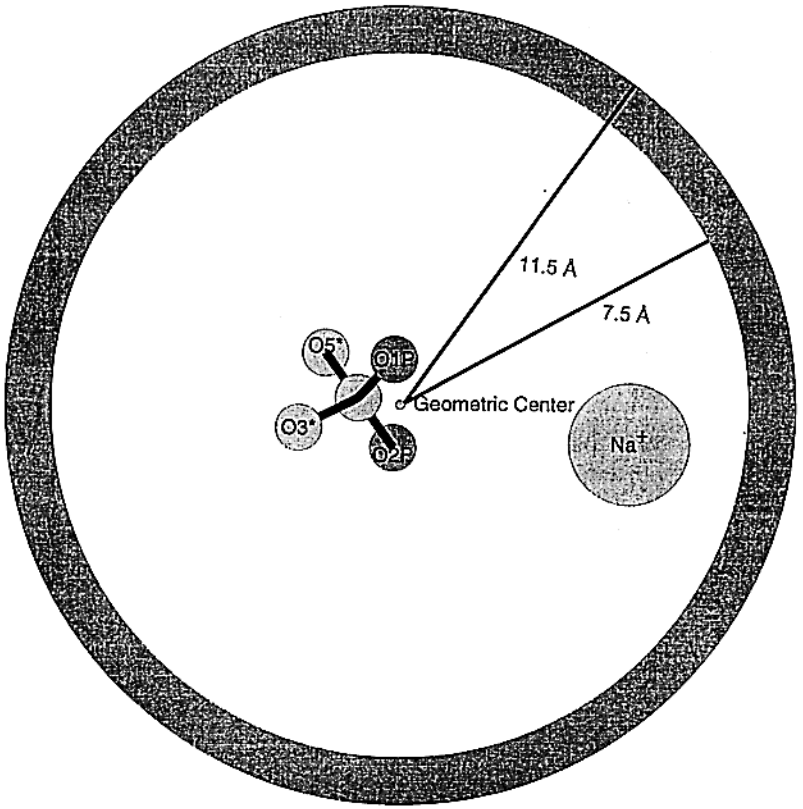


Figure 6 Neutral charge group constructed from backbone phosphate group and mobile counterion. The geometric center of the group is shown along with the switching region. The radii of atoms in the phosphate group are reduced in size for clarity. This arrangement shows the counterion at 6 Å from the P.

used along with 1927 SPC[72] waters in an HP box. The counterions were not grouped with the phosphates.

It is clear from Figure 8 that the simulation shows a rather strong peak at 8.25 Å, arising from the intrastrand P–P distances. This peak is an artifact; in canonical B–DNA the intrastrand P–P distance is 6.65 Å. Figure 8 reveals the switching function artifact discussed earlier: like charges separate beyond R_{off}, but unlike charges stay within R_{on}. The peak in the P–P distribution at 8.25 Å (with the distance between the geometrical centers of the group being at 8.5 Å) indicates that the adjacent phosphate groups repel each other until they are barely interacting. Once this has happened, they rarely sample distances less than 8.5 Å. One selected adjacent P–P distance as a function of time, shown in Figure 9, illustrates this point clearly. It also shows that this artifact does not

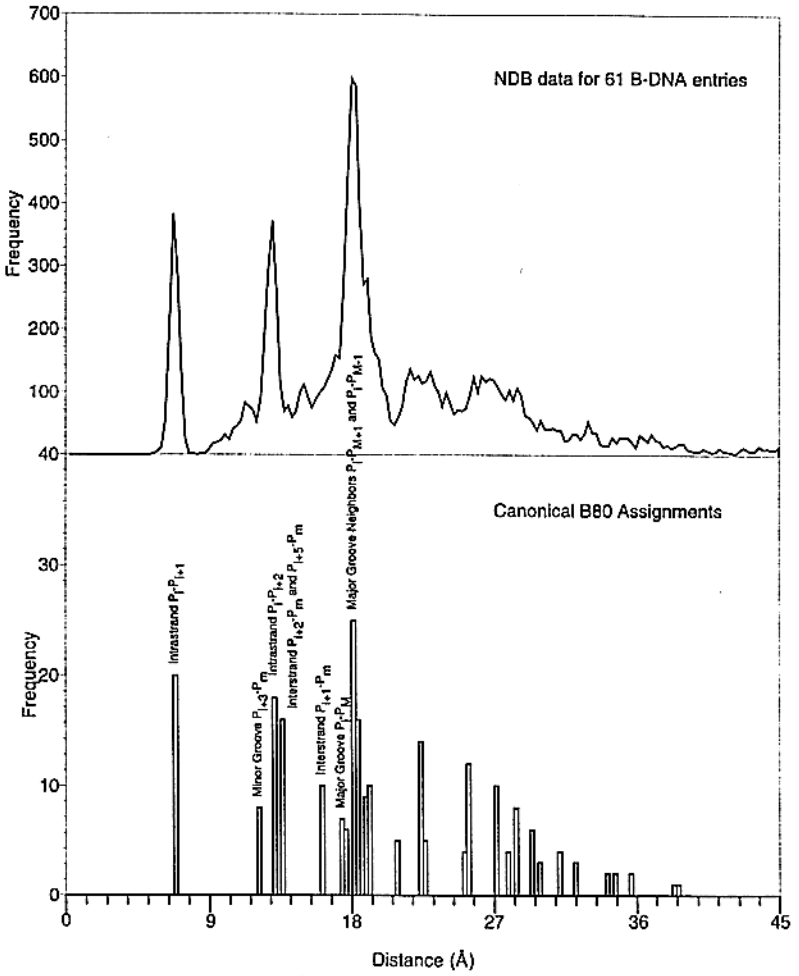


Figure 7 Distribution of phosphorus–phosphorus distances from 61 B-DNA X-ray crystal structures extracted from NDB (Ref. 1). *Top:* distribution constructed regardless of the base sequence or length of the oligonucleotides. *Bottom:* the P–P distribution for the canonical B form of DNA. The most interesting peaks are identified. Minor groove pairs are identified as P_{i+3}–P_m. Assuming that the molecule has an O5′ cap, $1 \le i \le N - 4$ and $m = 2N - i + 1$, where N is the number of residues per strand and m is the index of minor groove partner of phosphate i. Major groove pairs and their neighbors are identified also. For major groove pairs, $1 \le i \le N - 5$ and $M = 2N - i + 5$, where M is the index of the major groove. The unidentified peaks between the intrastrand second neighbors and major groove correspond to interstrand phosphates.

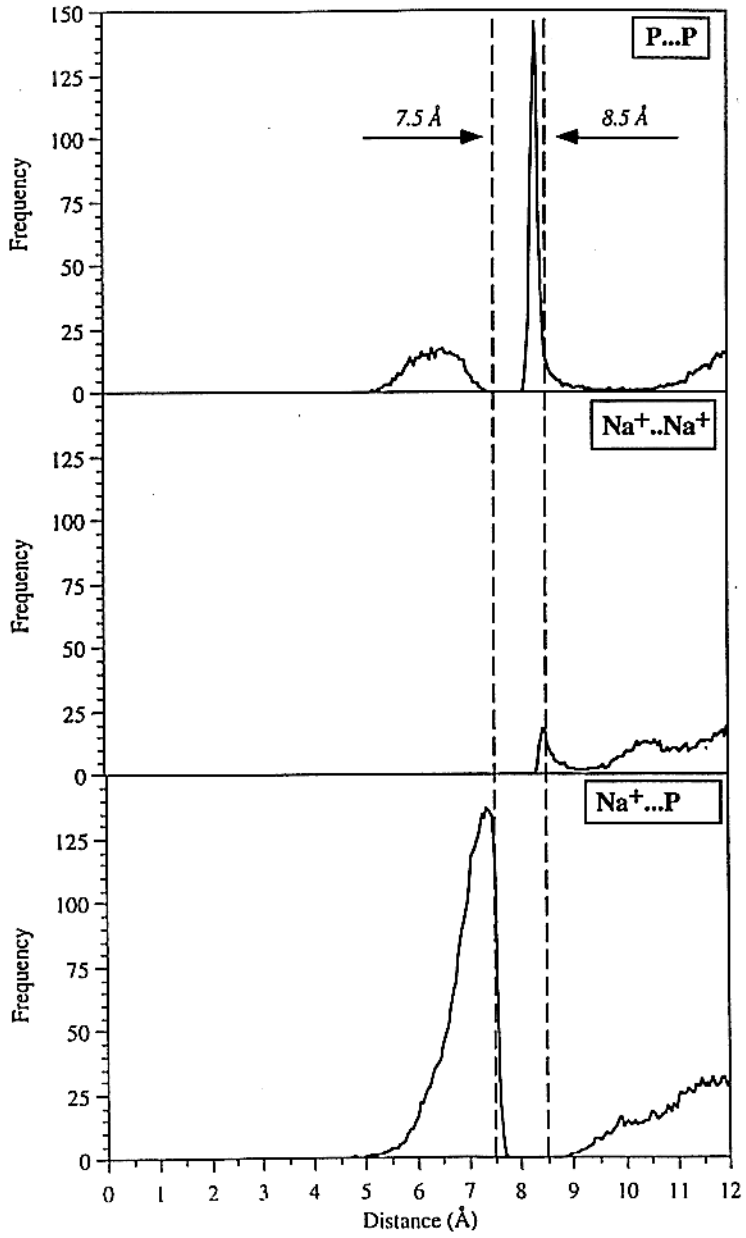


Figure 8. P–P, Na^+–Na^+, and Na^+–P radial distributions from a 500 ps simulation of d(CGCGAATTCGCG) using a 7.5–8.5 Å switching function. The counterions were not grouped to the backbone phosphate groups. Switching function artifacts arising from its short range cause the abnormal peaks seen in the figure. See text for detailed explanation.

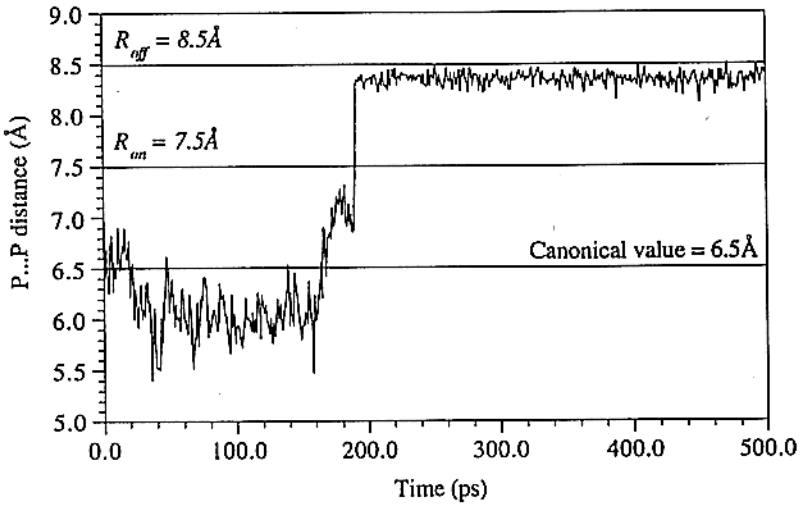


Figure 9 A selected P–P distance trajectory contributing to Figure 8. In the early part of the run, the P–P distances hover around the initial value. The separation due to repulsion happens at 200 ps, facilitated by the short switching function range. Once separated beyond the cutoff, the ions never return to interact together.

manifest itself until after 200 ps, suggesting that in short simulations such phosphate group repulsions may not be a factor to worry about.

We point out that the separation of adjacent phosphates also has an upper limit, dictated by the geometry of DNA. Changing the R_{off} to 11.5 Å has been shown to remove this artifact[81,82] because the adjacent phosphates can never separate to 11.5 Å due to covalent link restrictions. On the other hand, values in the 7.5–11.5 Å range introduce other problems, whereby minor groove phosphates interact with each other, but not the major groove phosphates.

Shifting Functions

Shifting functions[5,77] provide an alternative to smoothly truncated potential energies. As the name suggests, the true potential is "shifted" so as to make it zero at the cutoff distance. One form of a shifted potential is:

$$\mathrm{SF}(r, R_{\mathrm{off}}, \alpha, n) = \begin{cases} 1 - \alpha\left(\frac{r}{R_{\mathrm{off}}}\right)^n + (\alpha - 1)\left(\frac{r}{R_{\mathrm{off}}}\right)^{\alpha n/(\alpha-1)} & 0 < r < R_{\mathrm{off}} \\ 0 & r > R_{\mathrm{off}} \end{cases} \quad [5]$$

Here α and n are positive integers and are generally set to 2.[77] At short distances the electrostatic energy is reduced in comparison to the true potential. As a result, attraction between two oppositely charged ions is reduced, as is

repulsion between like-charged ions. Application of this function to systems using neutral charge groups is the same as for switching functions discussed earlier, and the methodological problems noted above are equally applicable. An alternative is to use a force-shifted function,[77] in which the force rather than the energy is required to smoothly drop to zero at the cutoff distance. This is more appropriate for MD simulations, where the trajectory generation is driven by the force.

Ewald Summation

Ewald summation[83] is claimed to be one of the most accurate methods for treating electrostatic interactions when PBCs are used.[84] In fact, the electrostatic interactions are calculated between all atoms located in the central box and between all atoms of the central box with their images in the neighboring boxes. The electrostatic interaction energy in a periodic system is given by[5]

$$V_{qq} = \frac{1}{2}\left[\sum_{|\mathbf{n}|=0}^{\infty}{}' \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \right] \qquad [6]$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, $\mathbf{n}$ is the lattice vector $\mathbf{n} = (n_x L, n_y L, n_z L)$, where L is the length of a cubic box and q_i and q_j are net atomic charges centered on atoms i and j. The prime appearing in the first summation indicates that the interaction for $i = j$ is omitted in the central cell, $|\mathbf{n}| = 0$.

Equation [6] is valid only for a cubic box, and extensions to rectangular boxes are achieved by modifying L. This is a conditionally convergent sum where the resulting energy is dependent on the order in which the double summation is carried out. A convenient mathematical transformation of Eq. [6] into two expressions with better convergence properties results in Eqs. [7] and [9].[5,66]

The first equation, in direct space, is

$$V_{\text{direct}} = \frac{1}{2}\left[\sum_{|\mathbf{n}|=0}^{\infty}{}' \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j \operatorname{erfc}(\alpha|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} \right] \qquad [7]$$

where erfc is the complementary error function, α is a parameter that controls relative convergence of direct space and reciprocal space summation, and all other values are as defined in Eq. [6]. The complementary error function is a continuous function of x and falls to zero with increasing x:

$$\operatorname{erfc}(x) = 1 - \frac{1}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2}\, dt \qquad [8]$$

The second equation is in reciprocal space, where $\mathbf{k} = 2\gamma\mathbf{n}/L$.[2]

$$V_{\text{reciprocal}} = \left[\frac{1}{\pi L^3} \sum_{k \neq 0} \frac{4\pi^2}{\alpha^2} q_i q_j\, e^{-k^2/4\alpha^2} \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) \right] \qquad [9]$$

Here, the parameter α controls the relative convergence of direct and reciprocal terms. Increasing the value of α causes the direct space sum to converge rapidly, which makes the reciprocal sum converge more slowly. The physical principles behind this reformulation[66] assume that a given point charge is surrounded by a charge distribution of equal magnitude but of opposite sign. The shape of the charge distribution, which acts like an ion atmosphere and screens interactions between the given point charge and the others, is assumed to be Gaussian shaped, with its width controlled by the parameter α.The screened interactions are evaluated in direct space and essentially determine the short-range interactions. A canceling sum, equal in magnitude to the screening distribution but of opposite sign, is added to the electrostatic potential to recover the potential due to the original set of charges. This term is summed in the reciprocal lattice space and can be interpreted to model the long-range interactions. The reciprocal space summation is computed by means of Fourier transforms[66] of the canceling distributions, and the sum is then converted back to real space.

It is interesting to note that a generalized formalism advanced by Berendsen[6] on the separation of electrostatics into short- and long-range terms results in the Ewald summation being a special case. The reformulation of Ewald summation implicitly contains a self-energy term of the canceling distribution with itself. This is corrected by subtracting the following term:

$$V_{\text{self}} = \frac{\alpha}{\pi^{1/2}} \sum_{i=1}^{N} q_i^2 + \frac{2\pi}{3L^3} \left| \sum_{i=1}^{N} q_i \mathbf{r}_i \right|^2 \qquad [10]$$

Ewald summation presented above calls for the calculation of N^2 terms for each of the periodic boxes, a computationally demanding requirement for large biomolecular systems. Recently, Darden et al.[85,86] proposed an N log N method, called particle mesh Ewald (PME), which incorporates a spherical cutoff R_c. This method uses lookup tables to calculate the direct space sum and its derivatives. The reciprocal sum is implemented by means of multidimensional piecewise interpolation methods, which permit the calculation of this sum and its first derivative at predefined grids with fast Fourier transform methods. The overhead for this calculation in comparison to Coulomb interactions ranges from 16 to 84% of computer time, depending on the reciprocal sum grid size and the order of polynomial used in calculating this sum.

In the PME method, now implemented in AMBER 4.1, the value for α is chosen to ensure that the direct space sum vanishes at the specified cutoff. The algorithm requires the user to specify an acceptably small tolerance for the

direct space sum. Then an iterative binary search for α is conducted, starting with $\alpha = 0.5$, until the $\mathrm{erfc}[(\alpha R_c)/R_c]$ is less than the tolerance at the assumed spherical cutoff distance R_c. The iterative binary search for α is done by first calculating the value of $\mathrm{erfc}[(\alpha R_c)/R_c]$ using the initial value of α and comparing it to the tolerance. Depending on whether the complementary error function is greater or less than the tolerance, the next search for α is directed toward the interval [0.5,1.0] or [0.0,0.5]. At each point, the value for α is chosen to be the midpoint of the interval, and this procedure is continued until the complementary error function is less than the tolerance.

The foregoing example assumes the range for α to be [0,1]. The calculated value of α for the Drew–Dickerson dodecamer with a fully charged phosphate backbone and equivalent number of counterions with 9 Å cutoff is 0.3483.[87] Energy conservation is used to assess the accuracy of Ewald summation relative to other methods. Several trial runs, during which α and the parameters for reciprocal summation are adjusted, may be required to reach desired levels of energy conservation. Superior energy conservation relative to cutoff-based methods has been demonstrated using the PME.

Ewald summation methods look very promising, but the implied periodicity of the system makes it more appropriate for crystal simulations than for solution simulations. No other method, however, allows accurate evaluation of the long-range electrostatic interactions required to treat the counterions in DNA simulations, and therefore the Ewald approach is worth exploring. In addition, it provides a means for carrying out concentration-dependent structural studies using additional ions from added salt. Because all solute and ion interactions are included with those in several neighboring cells, the approach cannot model infinitely dilute aqueous solutions, and this is a serious limitation. In addition, the method is very computer resource intensive. Detailed structural analyses of the DNA and counterions in simulations using Ewald summation are required to assess both the validity and the range of its applicability. A systematic study comparing results from various force fields and available methodologies, including the PME method, is under way in our laboratory.[88]

Restraints and Constraints

Restraints can be used to bias the computer simulations so that better sampling of a compact region of phase space is achieved in a short time period. *Constraints*, on the other hand, are typically used to freeze certain uninteresting degrees of freedom, including high frequency bond vibrations, thereby allowing a larger time step in an MD simulation. The difference between a restraint and a constraint is subtle. We use the following operational definition for distinguishing the two. Restraining functions are typically harmonic, with the target value of the restraint being the minimum and the shape of the harmonic function dictating the strength of the restraint. Using a harmonic

function to restrain the mobile counterions forces them to remain within a certain distance of the DNA surface, resulting in better sampling of the contact ion atmosphere than otherwise. Restraints are also used when phenomenological effects about the structure of the molecule are not included in the force field but need to be incorporated somehow. NMR restraints deployed to restrict the interproton distances to match experimentally measured two-dimensional NOE data represent another example of their use.[22] Constraints, on the other hand, require that selected degrees of freedom be fixed at their respective target values or held nearby with a very small tolerance. SHAKE is a commonly used algorithm to constraint any covalent bond with a hydrogen atom to its initial length, with a typical tolerance of 10^{-5} Å. SHAKE uses Lagrange's method of undetermined multipliers to simultaneously constrain bond lengths to their respective target values. Detailed descriptions of SHAKE and its implementation can be found in several sources.[23,66,67]

In MD, the addition of restraint energy calls for the evaluation of the corresponding forces, so the restraining function must be chosen with care. Typically, harmonic or hemiharmonic functions are used. Hemiharmonic restraints are essentially one side or the other of a symmetric harmonic function.[89] In addition, the restraining energy in the initial structure should be considered. If it is too high, these large forces may bring the pair of atoms to their restrained distance too quickly, causing severe problems in parts of the system that have not had sufficient time to respond to this sudden movement. The result is likely to be a system that is initially very unstable.

Hemiharmonic restraints for P–Na^+ distances are appropriate for restricting the motions of Na^+ ions around the surface of the DNA.[89] Force constants between 25 and 50 kcal/(mol Å^2) have been used; a restraint function beginning at a distance of 4.5 Å was found to keep the Na^+ ions within 6 Å of the associated phosphate. Restraints of this type have severe limitations however. For example, a counterion must be within 6 Å of its preassigned phosphate, and counterion exchange among the intrastrand phosphates is disallowed. However, the results are first-order approximations that still provide insight into the structure of DNA.

Restricting the ion movement with respect to the helical axis is more appropriate, but because the structure of DNA changes so much during the course of an MD run, the definition of "helix axis" itself becomes arbitrary. Helical axis based restraints are useful for MC simulations of canonical DNA, where the definition of a helical axis is straightforward, however.

The 5′ and 3′ base pairs of short DNA sequences used in computer simulations have a different environment because they are exposed to the solvent considerably more than the interior base pairs. These end base pairs tend to exhibit different structures and dynamical motions, commonly referred to as "fraying" or "end effects," in comparison to the interior ones. Such motions tend to break the Watson–Crick hydrogen bonds between the end bases, causing structural deformations in the neighboring base pairs. If the DNA segment

being modeled is short, this effect is transmitted to the entire structure, robbing the DNA of its double-helical character. This artifact arises from using a short length of DNA in the simulations. To overcome the loss of double helicity, a strong harmonic function [with force constants in the range 25–50 kcal/(mol Å^2)] can be used to restrain the base pair hydrogen bond distances to their canonical values. Alternatively, AMBER allows the use of atom coordinate constraints, referred to as the "belly" option, to hold (constrain) the atoms of end base pairs to their initial positions.

Validation and Analysis

Validating the trajectory for any computer simulation is essential. Validation involves computing properties that can be compared with those from experiment. It may also involve visual examination of the molecular structures, to search out any obvious abnormalities. Once the computer model has been validated, additional properties can be derived from the trajectory. This phase of study requires the availability of an exhaustive set of analysis tools. MD Toolchest,[24] developed at Wesleyan University, is such a collection, providing tools to extract both structural and energetic information from a simulation by a wide variety of MC and MD programs.

Analysis of DNA structure from MD simulations is complicated because DNA is very flexible. Qualitative analysis involves animating the MD trajectories for visualization. Quantitative analysis is done by monitoring DNA conformation indices, such as backbone torsions and helicoidal properties, and morphological indices, such as groove widths. The program "Curves, Dials and Windows,"[90,91] which is found in MD Toolchest, provides a means of analyzing DNA dynamics in an exhaustive fashion. Several articles describing the use of this tool have appeared in the literature.[89,91–95]

Morphological indices of the DNA double helix provide further insight about the structure of the nucleic acid molecule. The structure and dynamics of the minor and major grooves of DNA are extremely important. Many biological functions of a DNA, including drug and protein binding, are dependent on the groove structure. A quantitative description of the DNA groove structure is very difficult to provide except for the canonical DNA forms. Several attempts have been made to quantify the groove widths for irregular DNA,[24,34–36] but none proved satisfactory. MD Toolchest provides at least two tools to monitor groove width, one using the P–P distance across the grooves and the other called "unrolled helix." In the unrolled helix, the backbone atoms are projected onto a cylinder and the cylinder is cut open to make it flat. An ensemble average of backbone atom projections onto the cylinder then provides a plot showing the density of states for various groove widths.

DNA bending[37] is yet another index of interest. MD Toolchest provides two tools, "Bending Dials"[38] and "Persistence Analysis,"[96] to characterize bends in the DNA helical axis. Bending Dials calculates stepwise DNA bending

and presents the amount and direction of bending in the form of a dial. Persistence Analysis follows the theories developed to study the structure and flexibility of long polymers to quantify the extent of "straightness" in DNA and the junctions where it bends.

Another important component of the analysis phase, and which is especially germane to this chapter, involves assessing the structure of the ion distribution around DNA.[18,97] Ensemble averages of the cumulative fraction of the counterions as a function of distance from the global DNA helical axis allow one to estimate the fraction of "condensed" counterion. (Cumulative fraction is more convenient to compute than a concentration, which would necessitate calculation of the volume of a shell.) When the ions are restrained, this fraction and the concomitant structure will most likely be found to be a function of the restraint parameters.

Thermal ellipsoids are indicative not only of the extent of atom or group motion but also of their directionality. These ellipsoids can be overlaid on the average positions of ions and of backbone atoms to promote an understanding of the extent and cooperativity of backbone–counterion motions. Figure 10 shows the motion of counterions around DNA: successive locations of ions along the trajectory are connected, the initial ion positions are indicated by means of small spheres, and the average structure of DNA is displayed for

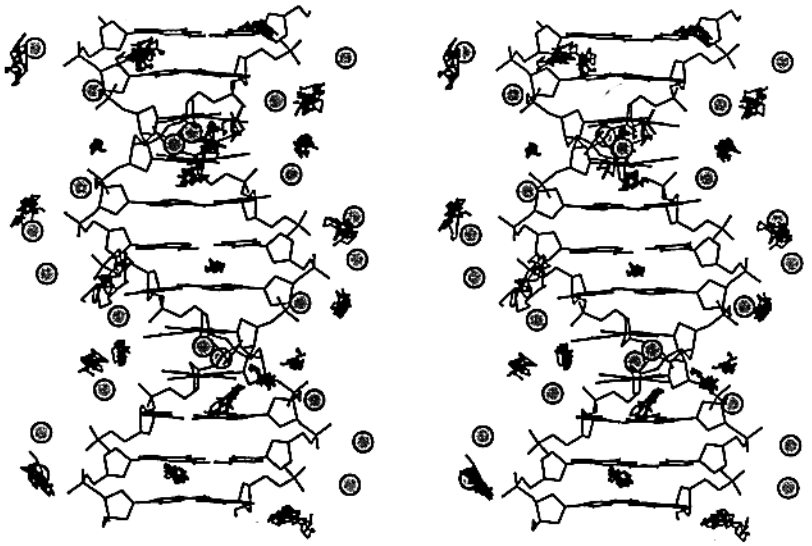


Figure 10 Stereo diagram of ion dynamics: the average structure of DNA from the last 100 ps of a 500 ps MD run on d(CCAACGTTGG) with explicit counterions and SPC waters. The initial positions of ions (at 400 ps) are shown as circles. The positions of the ions at every 2 ps are connected by lines.

clarity. Location of ions relative to the backbone phosphate, whether they occupy positions in the grooves or along the O1P–P–O2P bisector away from the DNA into the solvent, is a structural index providing clues about how best to place ions in the initial setup of future DNA simulations. Good initial geometries will help accelerate convergence in the simulation. Usually the interaction of counterions is assumed to be maximal with respect to the phosphate backbone. However, ions in grooves are in the proximity of the nucleotide bases, and examination of base–ion interactions can reveal information about their relative importance in counterion atmosphere around the DNA.

Hydration of DNA can be characterized by examining the solvent molecules in the grooves of DNA.[65,98,99] Minor grooves of DNA are generally found to contain water molecules occupying relatively fixed positions along the groove, giving rise to the so-called spine of hydration from MC simulations. Similar characterization from MD is more difficult to accomplish because the grooves are constantly changing and the definition of localized water becomes subjective. Nevertheless, groove hydration is an important structural index that should be examined. The proximity criterion[100] is a concept developed by our group to characterize hydration around nonspherical solutes. It is implemented as a computer program in MD Toolchest. The proximity criterion examines each water molecule around the solute and "assigns" it uniquely to the closest solute atom. This criterion can be used to examine the hydration around single atoms or functional groups in the molecule.

The pair correlation function $g(R)$[79,97] is often used to monitor the structure of ions in solution surrounding DNA. The $g(R)$ has its origins in liquid state theory and measures the local density fluctuations relative to the bulk density. Calculation of $g_{Na^+-Na^+}(R)$ reveals the preferred arrangements of Na^+ ions in water as a solvent-separated ion pair. Any simulation artifact is likely to show up at and beyond the cutoff distances, as demonstrated by Auffinger and Beveridge.[79] Similarly, solvation around Na^+ can be understood in terms of the $g_{Na^+-Owater}(R)$ depicted in Figure 11. The coordination numbers, reflecting the number of solvents within a specified distance of a given ion, can also be evaluated. It should be mentioned that $g(R)$ may not be the best index to compute because the spherical shells drawn around the ions when accumulating data for $g(R)$ also include some excluded volume of the DNA. The added material hinders $g(R)$ comparisons of counterions in DNA simulations with those from simulations of aqueous salt solutions. As a result, simple interionic distance distributions should be used for such direct comparisons.

Analysis of only selected parts of the system during the validation phase of research is a dangerous exercise and can lead to erroneous conclusions. We illustrate this by means of results from an unpublished 500 ps simulation on the Drew–Dickerson dodecamer sequence in its canonical form, surrounded by ~2000 SPC waters in a hexagonal box. In this simulation, each counterion was associated with a phosphate group to form a neutral charge group and was restrained by a hemiharmonic potential centered at 6 Å with a force constant of

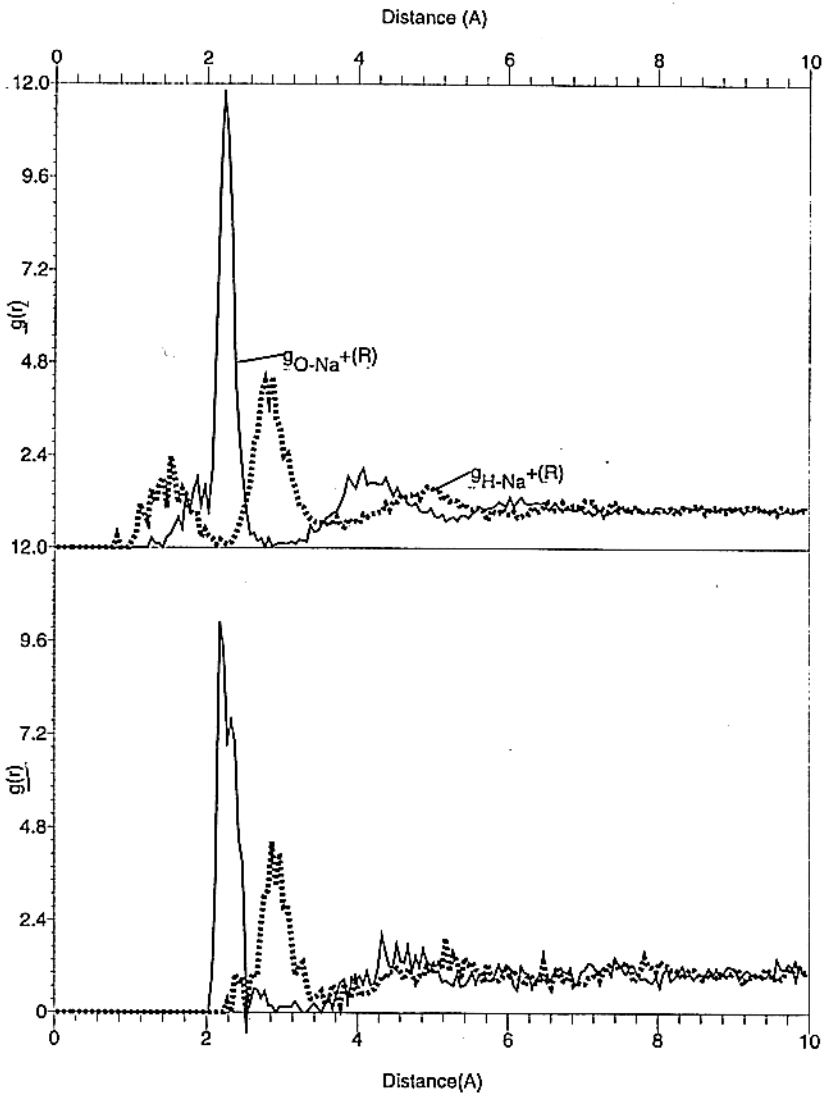


Figure 11 Ion–water pair correlation function $g(R)$. *Top:* the $g(R)$ for Na^+ and water oxygen and hydrogens in a simulation using the neutral charge group model having counterions coupled to backbone phosphate groups. The combination of short switching function range and the ion coupling causes an artifact that prevents nearby water molecules from seeing the Na^+. *Bottom:* changes in $g(R)$ are evidence that uncoupling the ion immediately corrects the problem.

25 kcal/(mol Å²). A group-centered switching function from 7.5 to 8.5 Å was used to truncate all nonbonded interactions. Equilibration of the waters and ions prior to the MD simulation took 3 million MC steps. Examination of the DNA and counterion structures from the MD simulation revealed no obvious abnormalities. However, the $g_{Na^+-O_{water}}(R)$ and $g_{Na^+-H_{water}}(R)$ showed that a distribution of hydrogen atoms from the water was closer to the monitored Na^+ ions than was the oxygen, as a result of a simulation artifact. Subsequent analysis revealed, in accordance with our earlier suggestion, that the source of the problem was the grouping of the sodium ions with their respective phosphate groups to achieve neutral charge groups. (Bound sodium atoms that are not grouped will not show this artifact.) Both the range of the switching function and the structure of the counterion around the DNA at 6 Å created a situation in which waters, at very short distances from the Na^+, were not interacting with the Na^+. Treating the sodium as a separate single-atom charge group immediately fixed this problem, as illustrated by the $g_{Na^+-O_{water}}(R)$ in Figure 11.

The stability of molecular dynamics simulations is another concern. Recently, Auffinger, Louise-May, and Westhof[68] studied the constancy of MD trajectories when slight perturbations (e.g., modifications of the initial randomly assigned velocity distribution) are introduced in the equilibration protocol. Such small perturbations had the effect of creating diverging MD trajectories. These investigators showed that the divergence calculated for a set of 10 trajectories, each starting from the same initial nuclear configuration but with different initial velocities drawn from Maxwellian distributions, was indicative of deficiencies in the simulation protocol employed. They proposed using a multiple MD strategy as a diagnostic for estimating the reliability of a set of trajectories and consequently of the underlying theoretical model.

ATOMISTIC COMPUTER SIMULATIONS: EXAMPLES

Monte Carlo Approaches

Monte Carlo calculations on simplified model systems representing the DNA, counterions, and water solvent have been carried out by several research groups. Le Bret and Zimm[101] reported two such calculations. The first used an impenetrable cylinder embedded with a linear array of charges to represent DNA backbone and the other used a double-helical charge array on an impenetrable cylinder. The mobile ions were treated as hard spheres, and the ionic interaction between the ion and the model DNA were modulated by the solvent, which was treated as a dielectric continuum with a dielectric constant of 80. Ion distributions around the cylinders were calculated and compared, but there were no significant differences between the two models, possibly because the average electrostatic interactions between counterions and two cylindrical models were similar. These authors also evaluated the effects of ion radius and ion charge on the ion distribution around the model cylinder. The condensed counterion concentration on the surface of the DNA was in general agreement with that predicted by Manning.[14]

Murthy, Bacquet, and Rossky[102] extended the work of Le Bret and Zimm[101] by modeling the various interactions in a more detailed fashion. The DNA–counterion interactions were evaluated with a softer, r^{-9} dependent repulsive potential rather than the hard sphere model used by Le Bret and Zimm. The electrostatic interactions between DNA and the ions were calculated using a special logarithmic function. The solvent was also treated as a dielectric continuum in this study. A potential to represent the long-range interactions between the ions and the images of the charge distribution in the adjacent cells along the helical axis of the DNA was included. Murthy, Bacquet, and Rossky calculated the counterion concentration at the surface of the DNA and found that it compared well with the Manning theory. They pointed out the importance of long-range electrostatic interactions by showing that its neglect can result in a 12–18% underestimation of the counterion concentration. The independence of counterion concentration as a function of the ionic strength of the medium was demonstrated, but it was shown that the Manning radius is reduced as the ionic strength is increased.

Mills, Paulsen, et al.[103–105] performed a systematic study along similar lines, exploring the structural correlations among the counterions, as well as the validity of Manning's theory. They found that the net positive charge in a fixed volume around the DNA indeed varied as a function of the ionic strength, a conclusion similar to that of Murthy et al.[102] Conrad, Troll, and Zimm[106] incorporated a dielectric discontinuity in the electrostatics in an attempt to understand its effect on the distribution of counterions in the DNA grooves. This discontinuity uses two different dielectric constants, one for interactions "inside" the DNA and another for the bulk solvent. Resulting ion distributions showed very low probabilities for ions in the grooves, especially the major groove. The authors ascribed this effect to the ion–ion repulsion and the low permittivity of the helix. Dielectric constants for models such as these can only be guessed at. Consequently ϵ values are difficult to establish and introduce a high level of arbitrariness into the model.

Detailed MC studies characterizing the minor groove hydration of a d(CGCGAATTCGCG) dodecamer sequence was reported by Subramanian, Ravishanker, and Beveridge.[65,99] This work focused on gaining additional perspective on the "spine of hydration,"[107] a network of crystallographically ordered water in the central adenine–thymine (AT)-rich region found in the X-ray crystal structure of this sequence. Periodic boundary conditions were applied to provide a solution environment. The DNA in its canonical B form was surrounded by 1777 TIP4P[71] water molecules and placed in a central hexagonal prism box. DNA–water interactions were treated under the mini-

mum image convention and were modeled using the AMBER force field.[108] A spherical cutoff of 7.5 Å was applied to all water–water interactions. Explicit counterions were omitted, but to achieve a system with a net charge of −0.24 per residue (Manning theory), net atomic charges of all DNA atoms were reduced equally. This was done instead of reducing just the backbone phosphate charges because the latter approach would have resulted in hydrophilic base atoms having significantly higher charges than the backbone phosphate atoms. The resulting configurations were analyzed, and the authors concluded that the spine of hydration was not specific to the AT-rich region but extends to the flanking cytosine–guanine (CG) regions also.

Jayaram et al.[18] performed a systematic study of the effects of electrostatic interactions on the counterion condensation around DNA. They used a 20-mer of electrically neutral sodium–DNA, with the DNA fixed in its canonical B form. The mobile counterions were placed randomly in a 50 Å radius cylinder around the DNA, and the solvent was modeled as a dielectric continuum. Four dielectric treatments, ranging from Coulombic interactions with constant dielectric to a dielectric saturation model with a modified Coulombic potential introducing dielectric discontinuity, were studied. The dielectric saturation model used a modified Hingerty sigmoidal function[109,110]:

$$\epsilon(R) = D - \frac{D-1}{2}\left[(RS)^2 + 2RS + 2\right]\exp(-RS) \qquad [11]$$

Here, $\epsilon(R)$ is the distance-dependent dielectric constant, D is the dielectric constant plateau value at long distances, and S is the slope at the sigmoidal segment of the function. The parameters for this function were chosen such that the dielectric constant at distances less than 3 Å becomes 20 but at distances greater than 15 Å becomes 80. The function increases sigmoidally in between, with a value of 50 at 5 Å. The authors concluded that independent of the dielectric model, counterion condensation was generally consistent with Manning's counterion condensation theory. However, the structure and the energetics of the ions around DNA differ significantly among the models. They also studied the effects of added salt on the counterion atmosphere and found that the dielectric saturation model was the only one maintaining ionic strength independence, as predicted by Manning.

Gordon and Goldman[111] carried out Monte Carlo simulations on a system consisting of a cylinder having uniform charge distribution as well as a helical distribution, surrounded by 15 counterions and explicit SPC[112] waters. They noticed that the uniformly charged model significantly polarized the solvent water, causing the counterions to avoid regions proximal to the DNA. On the other hand, the helical lattice model produced results consistent with other studies[16,18] in which continuum solvents had been used.

Mills, Rashid, and James[113] recently reported detailed MC calculations on the ion distributions around A, B, and wrinkled D conformations of DNA. Their calculations were performed on the duplex DNA sequence d(AT-ATATATAT). The DNA atoms were assigned partial charges from the AMBER force field,[108] and hard sphere radii were used for all atoms in the DNA.[114] The univalent counterions and the negatively charged co-ions were treated as fully charged hard spheres of radii 3.0 Å, corresponding to the size of a hydrated Na^+. The solvent was treated as a continuum dielectric with dielectric constant of 80. It was found that the concentration of counterions in the major groove of A-DNA is approximately three times that for B-DNA. The D-DNA grooves are not sterically accessible to the ions and, accordingly, the simulation showed no counterion density in its grooves. Whereas conformation-dependent local fluctuations existed in ion distribution, the counterion concentration within a 24 Å diameter cylinder surrounding the DNA was conformation invariant. This finding is inconsistent with Manning's theory, according to which the condensed counterion fraction should differ between A and B forms arising from differences in the average phosphate separation along the helical axis. Mills, Rashid, and James also demonstrated the invariance of condensed counterion fraction to added salt within that cylinder.[113]

The MC simulations discussed above all used the canonical (T, V, N) ensemble. Grand canonical Monte Carlo (GCMC) simulations offer a powerful means of assessing the effects of ionic activity coefficients on the counterion atmosphere of DNA. The grand canonical ensemble is a constant (T, V, μ) ensemble, where the chemical potential μ can be transformed to mean ionic activities in polyelectrolyte solutions, and the system can be studied at fixed ionic activity coefficients. Several studies along these lines have been performed on a range of systems, from an infinitely long charged cylinder[115] to canonical DNA with explicit ions.[116] Detailed discussion of the GCMC simulation can be found in the original articles and references therein.[116]

Molecular Dynamics Approaches

Over the past 10 years, advances in computer technology have led to significant improvements in molecular dynamics simulations of biological molecules in terms of both system size and length of the simulation that can be handled. Nanosecond simulations of DNA[82,87,117,118] with explicit waters and counterions are currently viable. Yet despite such advances, there exist only a few MD simulations on DNA and even fewer simulations of DNA with explicit counterions. Most of these simulations treat the counterions implicitly, by systematically reducing the backbone phosphate group charges to account for the Manning counterion condensation theory. Although this approach produces a stable simulation, the system so modeled is not electrically neutral and therefore does not comprise a realistic physical description of the actual system. Many MD calculations, restrained using NMR data, also exist but are outside the scope of this chapter and can be found elsewhere.[22,41] Likewise, several other interesting classes of DNA, including triple helices,[118] and complexes of DNA,[119,120] have been the subject of MD studies exploring the treatment of

counterion and electrostatics. Here we review only studies applied to double-helical DNA. For additional reviews of DNA simulations from various viewpoints, the reader is referred to other sources.[3,121,122]

Singh, Weiner, and Kollman[123] reported MD simulation studies on a DNA pentamer sequence, d(CGCGA). They studied two variations of the DNA: one in the DNA backbone was fully charged, and in the other hydrated counterions ("*solvatons*"), emulating a hydrated Na^+, were used. Both mass and radius of the solvatons were adjusted to represent a hexahydrated Na^+. These studies used the AMBER force field with a nonbonded cutoff of 12 Å, and the electrostatic interactions were modulated by means of a distance-dependent dielectric function. The structural parameters of the two DNA models during the course of this 83 ps simulation were the found to be essentially the same, with the root-mean-square (rms) deviations for atomic motions about their mean positions in the range of 1 Å. Interesting correlations in the structural parameters of the DNA were discussed, and the calculated average values for the DNA twist and tilt angles agreed fairly well with the crystal data for the Drew–Dickerson dodecamer sequence. The solvaton model was found to have 10 base pairs per turn, more in line with canonical B-DNA than was the anionic model, which gave only 9 base pairs per turn. The number of base pairs per turn is extrapolated from the twist parameters of the pentamer. One solvaton migrated into the minor groove in the last 10 ps, but all others remained in the vicinity of the DNA backbone. The solvaton model for ions is appropriate in cases of interactions between DNA backbone and ions that are modulated by an intervening water. The differences in DNA structure resulting from an in vacuo simulation using simple counterions and solvatons have not been catalogued.

Kollman's group[124] repeated an earlier study on the same pentamer sequence, d(CGCGA), using explicit water molecules and eight explicit Na^+ counterions. This 106 ps study used 830 TIP3P waters in the form of a droplet around the DNA, and the authors chose a nonbonded cutoff radius of 10 Å for all atoms. The general structural features of the DNA, measured by average conformational indices, were found to be in the same range as those found in the authors' earlier study[123] using an implicit water model. The explicit water simulation damped phosphate motions, and 70–80% of the sugars were found to be in the C2′-endo conformation, with the rest in the C3′-endo form. Only two of the counterions, initially placed at the contact distance of 3.1 Å, remained near phosphate groups. At least one counterion diffused to the edge of the droplet, and one migrated into the minor groove region. Whereas inclusion of explicit water molecules increases the computational demand tremendously, there are several advantages to using explicit solvent models instead of implicit solvent models. For example, it is possible to explore specific hydrogen bonding interactions between the DNA and water molecules, along with the hydration of counterions, and these can be correlated with any available experimental data (e.g., water positions derived by X-ray or neutron diffraction studies).[1]

The atomic motions of solute atoms can be very different in explicit and implicit water models, resulting in significantly different MD trajectories.[94] In the study from Kollman's group,[124] the explicit water model did not change the DNA structure significantly from that obtained by means of an implicit water model, but the time scale of this simulation was too short to bring out such differences.

Van Gunsteren et al.[125] reported results from an 80 ps MD simulation on the octamer duplex d(CGCAACGC) including 14 Na^+ and 1231 SPC[112] waters; the GROMOS[126] force field was used. A solution environment was provided by applying periodic boundary conditions. A twin-range cutoff at 8 Å ensured that all interactions involving atoms at distances of less than 8 Å were evaluated at every MD step, but those at distances exceeding 8 Å were evaluated at every 10 steps, to conserve computer time. Na^+ ions were placed initially at positions having the greatest electrostatic potential around the DNA. The rms deviation between that simulated DNA structure and the canonical B-DNA was 2.2 Å, and the rms deviation with respect to the canonical A form was 3.5 Å. Two-dimensional NMR experiments on this sequence were available at the time the simulation was carried out, and 80% of the calculated interproton distances agreed with the experimental results. All explicit counterions were found in their solvent-separated states around the DNA, and no contact DNA–Na^+ pairs were found.

MD studies on B and Z forms of DNA with explicit waters and explicit counterions were reported by Swamy and Clementi.[127] G-C and A-T decamer sequences in their B forms were surrounded by a rectangular box with 1500 water molecules and 20 K^+ ions. In addition, a G-C dodecamer in its Z form with 1851 water molecules and 24 K^+ ions was studied. Water molecules in these studies were four-centered MCY[128] waters. These simulations were carried out for a total of 7 ps, with the first 3 ps serving as an equilibration period. The DNA in all cases was rigid, and only the ions and the waters were allowed to execute motions. The dynamical behavior of those ions showed them to be strongly bound to the DNA with restricted mobilities, a conclusion different from what counterion condensation theory and other simulations tend to suggest. The exploration of space by the counterions around DNA in the short time scale of this study was insufficient to permit the derivation of general conclusions about the ion mobilities, however.

Laaksonen et al.[129] reported a 70 ps MD simulation on poly(dG·dC) in its canonical Z form including explicit waters and K^+ counterions. The dodecamer sequence was capped at the ends with phosphate groups and required a total of 24 counterions. The DNA was placed in a cubic box and hydrated with 2279 SPC[112] water molecules. The DNA–K^+ and K^+–K^+ interactions were modeled with the Clementi and Corongiu potential,[130] and DNA–water interactions were calculated by means of the AMBER force field. A 10–12 hydrogen–bond potential was used to maintain the Watson–Crick hydrogen bonds. There were three different treatments of long-range electrostatics. The

first, using a shifting function cutoff at 10 Å, showed continuous temperature drift and large potential energy fluctuations. The second, a twin-range cutoff with all interactions within 10 Å updated every step and those between 10 and 15 Å updated every tenth step, improved the temperature stability. The third treatment, application of Ewald summation,[83] provided the best choice for both energy and temperature stabilities. Ewald summation, as described earlier, is a methodology used in calculating long-range electrostatic interactions by means of large numbers of periodic boxes. A 10 Å cutoff for evaluating the direct term in the Ewald summation was found to be optimal. The resulting configurations were analyzed for both DNA hydration and counterion structure. The counterions were shown to coordinate with the nucleotide bases, however, rather than the backbone phosphate. Compared to simulations of DNA-free salt solutions, the ions showed mobility reduced by about one-third. Detailed structural analysis of the DNA from this study was reported later by Eriksson and Laaksonen.[92]

Rao and Kollman[131] and Srinivasan, Withka, and Beveridge,[93] reported their results from MD simulations of an in vacuo model of the Drew–Dickerson dodecamer sequence. Both studies used the AMBER force field and Na^+ solvatons. The structure of the DNA during a 100 ps simulation time period converged to a form intermediate between the canonical A and B forms. Although the base pairs were symmetrically oriented with respect to the helical axis, resembling a B-DNA, the helicoidal parameter inclination had values characteristic for a canonical A form. Srinivasan et al.[93] found that initiating the trajectory from either the canonical A or B form of DNA resulted in structures for DNA that were practically indistinguishable. The solvatons were placed 6 Å from the O1P–P–O2P bisectors and were not restrained. During energy minimization, the solvatons remained within 6 Å of the phosphate groups, but at the end of the MD run they had moved far from the DNA. One hydrated counterion remained within 7.5 Å of its initial phosphate, and six others readjusted their positions to be near other phosphate groups. The remainder diffused from the DNA, and no solvatons were found in the grooves as a consequence of the large solvaton radius.

Zielenski and Shibata[132] performed a 60 ps MD calculation using the GROMOS force field[126] on the hexamer sequence ($dG_6 \cdot dC_6$) surrounded by 232 SPC waters and 10 octahedrally coordinated Na^+ ions. The 292 water molecules including the 6 water molecules per Na^+ were selected to mimic the first hydration shell around the DNA as a 5 Å layer. The counterions were placed to ensure that water molecules from the hydration shell of the ion formed bifurcated hydrogen bonds with adjacent phosphate oxygens. The SHAKE algorithm was used to restrain distances for end base pairs to their initial values, to keep them from fraying.

Several interesting features regarding the structure of the DNA were observed during the course of this simulation. The DNA hexamer retained an overall B conformation throughout Zielenski and Shibata's simulation, and the

authors observed a high propeller twist and narrowing of minor groove. These results are consistent with results from the first 60 ps of other longer simulations using the GROMOS force field and SPC waters.[81,133,134] However, the longer simulations show that the initial structures do not persist, and the DNA structure changes greatly after 100 ps. In a detailed energy component analysis, the authors found that the DNA–Na^+ and Na^+–Na^+ interactions were the major contributors to the overall potential energy profile, showing that the electrostatic interactions dominate the total energy. Finally, the authors noticed that two Na^+ ions were within ~3 Å of each other around 57 ps, and it was rare for this distance to drop below 5 Å during the simulation. Though such short distances of separation for like-charged ions are not intuitive, they have been found in several other instances, discussed later. This phenomenon corresponds to a "water-separated ion pair,"[97] discussed below. The configuration can be energetically favorable depending on the relative magnitudes of ion–solvent attractive energy and ion–ion repulsive energy. The authors followed up this study with another MD simulation[135] to probe the implications of GT mismatch in poly(dG·dC) sequences.

In a short 40 ps MD study of the decamer sequence d(CCAACGTTGG), Dickerson and co-workers[98] used the AMBER force field to characterize minor groove hydration. An X-ray crystal structure of this sequence provided input coordinates, and a 4.8 Å solvent bath around the DNA was constructed, resulting in a total of 491 waters. The DNA atoms were positionally restrained to their initial configuration, and a nonbonded cutoff of 8.5 Å was selected. No counterions were used in the simulation, but the charges on the O1P and O2P atoms for each phosphate group were reduced to obtain a net charge per residue of −0.3, to account for the condensed fraction of counterions. The resulting hydration patterns in the minor groove were correlated with the minor groove width. The narrow minor groove region spontaneously formed the "spine of hydration," defined by one first-shell water per base pair, whereas the wider groove showed individual base hydration. From this study,[98] the authors proposed that the spine of hydration is related simply to the groove width rather than to base sequence.

An MD simulation focusing on the structure and dynamics of both water molecules and counterions around the canonical B-form duplex of d(CGCGCGCG) was reported by Forester and McDonald.[97] These authors used nonbonded interaction parameters from various sources: AMBER for DNA atoms, Chandrasekhar et al.[136] parameters for Na^+ and Cl^-, and their own parameters for Ca^{2+} ions. Ca^{2+} parameters were derived by a trial-and-error method in which peaks in ion–water radial distributions calculated from MC simulations by means of assumed parameters were compared against experimental values.[137] This procedure was repeated adjusting the nonbonded parameters for Ca^{2+} until the two radial distribution functions agreed reasonably well. Five different simulations were carried out: one on fully charged, polyanionic DNA in pure SPC[112] water and four on electrically neutral systems

containing different combinations of SPC water and Na^+, Ca^{2+}, and Cl^- ions. The cations showed a strong preference for solvent-separated associations with DNA and little inclination for direct site binding, surprisingly irrespective of their valencies. The ions diffused somewhat into the region of the second hydration shell of DNA, leaving the number of first-shell waters essentially unaltered but significantly affecting their orientational ordering. The radial distribution function for Na^+–Na^+ pairs in the Na–DNA system interestingly starts at about 3 Å and has a maximum at 3.6 Å. This result was attributed to a "water-separated ion pair" arising primarily from pairs of ions coordinated to two different oxygens, and in the presence of added salt (NaCl), to the same oxygen on a given phosphate group.

Swaminathan, Ravishanker, and Beveridge[89] reported results from a series of MD simulations of the Drew–Dickerson dodecamer sequence obtained by means of the GROMOS87 force field. In vacuo simulations using solvated counterions (net charge 0.25) along with reduced charge phosphate (net charge −0.25) showed severe structural deformations within a few picoseconds of the simulation. The dodecamer in its canonical B form with the counterions placed along the O1P–P–O2P bisector at 6 Å was hydrated with 1927 SPC waters in a hexagonal prism box. All nonbonded interactions were treated with a switching function in the range 7.5 to 8.5 Å. Extensive MC simulations were required to equilibrate the system before stable MD trajectories could be generated. In the initial stages of the MC simulations, the ions were held fixed along with the DNA, and only the solvent molecules were allowed to move. Once the energy of the system had stabilized, the ions were included in the MC moves. The ions were found to settle at positions between two adjacent intrastrand phosphates during the course of the MC simulations. MD simulation of this model showed that the Watson–Crick (WC) base pairing was not conserved overall. A harmonic restraint [force constant of 5.0 kcal/(mol/Å^2)] was subsequently required to keep these hydrogen bonds from breaking, and a 140 ps trajectory was then generated. This model, referred to as the WC5.0 model, produced a stable trajectory, with the DNA retaining its B form and having several interesting local conformational features. Helical axis bending along with propeller twisting was shown to be consistent with the X-ray crystal structure. Comparison of NMR properties calculated for the collection of MD structures from this study showed excellent agreement with available experimental data.

We have subsequently carried out several other MD simulations on DNA both from the methodological and structural analysis points of view.[81,133,134] During the course of these investigations, it became clear that existing methods for treating long-range electrostatics, with the exception of Ewald summation, which was not implemented in our MD programs, did not work satisfactorily for simulations extending into the nanosecond regime. The problem was compounded in cases of explicit counterion simulations. We carried out a series of exploratory studies to document the deficiencies of several available truncation schemes, some of which are presented as illustrative examples in this review. In the meantime, a set of protocols was developed to carry out structural studies using the reduced charge model for phosphate.[82]

Fritsch and Westhof[138] carried out a series of 50 ps MD simulations on poly(dA·dT) using different dielectric models to modulate the electrostatic interactions. This study made use of the four different initial structures for DNA (proposed from earlier studies), without explicit counterions or water. The AMBER force field was used to describe all interactions, but the electrostatic interactions were modulated by a distance-dependent, as well as a sigmoidal, dielectric function developed by Ramstein and Lavery.[110] This study demonstrated that the DNA structure derived by using the sigmoidal dielectric function compares well with structures proposed from various spectroscopic data. The authors went on to characterize the behavior of hydrogen bonds in this system by evaluating lifetimes of three-center hydrogen bonds.[139] Fritsch and Westhof[140] applied similar techniques to the study of the conformational response of a hexamer sequence d(CGCGCGCG) in its Z form to modification of the G_4 residue by, *N*-2-fluorenylacetamide, a carcinogen.

Venable et al.[141] used the CHARMM force field to perform a series of MD simulations on the sequence d(CGCGATTCGCG), where T_6 is a TT mismatch. The effects of initial structure and atom velocities on the resulting structure were studied by performing several short 1 ps simulations. The low energy conformation from this screening was then used to perform 100 ps vacuum dynamics. This sequence was also studied with respect to 500 ps dynamics including explicit waters, modeled by a modified TIP3P potential. In all cases, there were no explicit counterions, and the backbone phosphate charges were adjusted to −0.32 to account for counterion condensation. The mismatch region was found to have the largest flexibility in comparison to the other Watson–Crick base pairs. The predominant conformations of DNA showed a large negative propeller twist at the mismatched base pair step.

Miaskiewicz, Osman, and Weinstein[142] reported results, obtained with the AMBER force field, from a 150 ps MD simulation on the Drew–Dickerson dodecamer sequence with explicit water molecules and counterions. This simulation was the longest DNA simulation reported before 1992. The system consisted of DNA in its canonical B form, 22 Na^+ counterions placed at 5 Å from the O1P–P–O2P bisector, and a 9 Å shell of waters around the DNA, resulting in a total of 1431 TIP3P waters. No periodic boundary conditions were employed; thus the system simulated a DNA droplet. The authors noticed that the DNA adopted two different conformations, one in the 20–60 ps range and the other at 100–150 ps. The structure of DNA in the latter portion of the simulation showed pronounced kinks near C3 and C9, and the double helix was severely underwound near the central AATT region. The counterions were found to be very mobile, and average counterion distributions showed approximately half of them in direct coordination with a phosphate oxygen, while the other half were at solvent-separated distances. The authors also characterized the hydration of base pairs and found the average coordination numbers for

GC and AT pairs to be 21.38 and 20.45 respectively. The possible inadequacy of the MD simulation length was discussed in the context of the convergence of the structure of DNA. The authors concluded that additional long simulations using varying simulation protocols and force fields are required to characterize the convergence of DNA structure.

The effects of explicit counterions and explicit waters on the structure of poly (dA·dT), compared to an implicit solvent model, were detailed in an MD study by Fritsch et al.[94] The Lipanov and Chuprina structure[143] for DNA was surrounded by a total of 18 ammonium counterions placed at 5 Å along the O1P–P–O2P bisector. This system was hydrated, resulting in a rectangular box containing 4109 TIP3P waters. The 50 ps MD simulation obtained by applying the AMBER force field to this system showed, after appropriate equilibration, lack of convergence and the need for longer runs. A comparison of the DNA structure in explicit versus implicit solvent treatments with experimental data revealed much better performance by the latter model. The authors also noted that the MD structures using the implicit water depended on the statistical ensemble used to model the system; a microcanonical ensemble performed better than the canonical ensemble. The simulation protocols used for canonical and microcanonical ensembles are different; therefore, the trajectories produced from the same initial structure for these ensembles will be different. The observation that the microcanonical ensemble performed better in this case simply means that it is traversing a more stable trajectory during the short simulation time. A different choice of simulation parameter might produce a different conclusion.

In general, it is advisable to run *(N, P, T)* ensemble simulations, because these are closest to most laboratory experimental conditions, making it easy to compare simulation results to available experimental data. A "spine of hydration," a string of localized waters along the minor groove of the DNA proposed by Drew and Dickerson,[107] was found in the explicit water simulation. The counterion dynamics showed that seven of the counterions settled into the major groove, eight into the minor groove, one near a phosphate group, and two relatively far from the DNA surface.[94] The counterions were found to lie in between phosphate groups, maximizing the favorable interactions, and many of the counterions were found at solvent-separated distances. This behavior is predominant in almost all explicit counterion simulations published in the literature so far.

Falsafi and Reich[95] used the AMBER force field to carry out implicit water simulations on the dodecamer sequence d(CGCGAATTCGCG) and two 14-mer sequences, d(GGCGGAATTGGCGG) and d(GGCGAAATTCGCGG) (referred to as A5 and A1, respectively). Hydrated counterions were placed 6 Å from the O1P–P–O2P bisectors and were constrained by means of a harmonic potential. A simulation of 89 ps on the dodecamer, starting from both the canonical B form and the X-ray crystal structure, revealed that the localized structural parameters fails to converge in this time period. The A1 and A5 sequences in their canonical B forms were then subjected to 200 ps and 1 ns simulations, respectively. Both simulations showed convergence of most structural parameters occurring within the 100–200 ps time scale. The authors recommended 200 ps implicit water simulation for comparative studies of DNA sequences. Subsequent explicit water simulations have shown that the assumptions on which this recommendation was based may be too simplistic.[82,87] Detailed structural analysis of the various DNAs were presented and compared, but no structural analysis of the counterions was considered.

We reported the first one-nanosecond MD study[82] on the dodecamer d(CGCGAATTCGCG). Net atomic charges of backbone phosphate atoms were scaled by 0.25 to account for counterion condensation,[14] and no explicit counterions were added. The system included a total of 2275 SPC waters[112] placed around the DNA in a hexagonal prism cell. Periodic boundary conditions were applied to simulate a dilute aqueous solution of DNA. Long-range electrostatic interactions were treated with a switching function from 7.5 to 11.5 Å. Switching functions used to truncate electrostatic interactions smoothly from R_{on} (7.5 Å) to R_{off} (11.5 Å) were discussed in detail earlier in this chapter. The trajectory was analyzed by means of a two-dimensional rms map, a tool used in detecting conformationally proximal structures called microstates, to reveal whether the DNA spanned two different microstates in the course of the simulation. The first microstate had an rms of 4.5 Å from the canonical B form and the second around 7 Å. Detailed structural analyses were carried out to document the variations in the structural parameters of DNA.

Recently, Cheatham et al.[87] reported results from a nanosecond dynamics simulation of DNA using both the spherical cutoff and Ewald summation; most stable trajectories were observed by applying the Ewald summation. The Drew–Dickerson crystal structure of the dodecamer d(CGCGAATTCGCG) with a fully charged phosphate backbone and an equivalent number of counterions surrounded by TIP3P waters in a rectangular box formed the initial structure. All interactions were modeled with AMBER 4.1,[69] using the parameter set of Cornell et al.[74] The long-range electrostatics treatment featured a charge-group based 9 Å spherical truncation cutoff (CUT), CUT in combination with complete evaluation of all solute–solute interactions (CUTSS), and particle mesh Ewald summation (PME).[85] Charge groups are functional groups or entire residues of a solute molecule. Implications of charge-grouping in interaction energy evaluations are discussed in detail below. A grid size of 1 Å was used in PME summation. The stability of the simulation was assessed by comparing the rms deviation of the structures from the MD trajectory to the initial DNA structure. CUT and CUTSS simulations showed a significantly diverging deviation, whereas PME structures showed a plateau with a 3.2 Å rms deviation after 200 ps. Similar behavior was demonstrated for an RNA and a protein. Whereas convergence criteria based on rms values are encouraging, detailed structural analyses of the DNA must be performed to fully assess the PME method.

Recently, York et al.[117] reported a crystal lattice simulation of 2.2 ns duration on the Drew–Dickerson crystal structure with fully charged phosphate backbone, explicit waters, and counterions. The unit cell (space group $P2_12_12_1$) contained four DNA duplexes, 1606 TIP3P water molecules, and 88 sodium ions. Ewald summation was used to calculate the long-range electrostatic interactions. Helicoidal and morphological properties of the DNA were found to be very close to the X-ray crystal structure on the nanosecond time scale. The rms deviation of the time-averaged simulation structure compared to the crystal structure was 1.16 Å. Isotropic B factors calculated from the simulation were in good agreement with the crystallographic values.[29] This study demonstrated that a proper treatment of long-range electrostatics, in this case using Ewald summation, is necessary to derive stable DNA structures.

In 1995 we completed a detailed comparative study of the Drew–Dickerson sequence with 22 Na^+ counterions and about 4000 TIP3P water molecules.[88] This examined the effects of initial ion configuration on the DNA structure as well as on the resulting counterion structure around the DNA. Three different simulations, each with a different starting position, were performed using AMBER 4.1[69] software and the parameter set of Cornell et al.[74] The initial ion configuration around the canonical B-DNA for the first simulation was generated using the "cion" module from AMBER 4.1, which places ions around the DNA at grid points having the greatest negative electrostatic potentials. The length of this first simulation was 1 ns. The ion configuration for the second MD simulation was generated by performing a counterion MC[18] simulation with implicit solvent model using the Lavery sigmoidal dielectric function[110] and selecting the minimum energy configuration of ions around the DNA as starting points for MD. In counterion MC simulations, ions are moved around rigid DNA to sample energetically favorable configurations. The third simulation was carried out beginning with ions placed along the bisector of O1P–P–O2P, 6 Å from P. The length of each MD trajectory for the last two simulations was 300 ps. Periodic boundary conditions were applied and the long-range electrostatic interactions calculated by means of the PME[85] method implemented in AMBER 4.1. Finally, a 200 ps simulation on the same system was carried out, using cion to create the initial configuration but with a twin-range cutoff from 7.5 to 11.5 Å employed in calculating the electrostatic interactions.

All simulations using PME showed that the DNA structure remained in its canonical B form. The conformational and helicoidal parameters compared with all B-DNA structures available in NDB showed striking similarities. An interesting observation concerning the counterions was made. Considerable counterion mobility was noticed in all simulations, with proximal ions on the DNA surface diffusing toward the edge of the simulation box, while distal ions far from the DNA began approaching its surface. An average counterion density of 0.76 within a 17 Å radius from the DNA helical axis is consistent with Manning's counterion condensation theory.[14] Comparison of rms deviations of DNA structures to the canonical B-DNA for the three simulations clearly shows that the PME method produces a description of DNA structure and dynamics far superior to those derived from a twin-range cutoff. Our ongoing analysis on the DNA–ion and ion–ion structural correlation is likely to bring more insight into Ewald summation methodology.

CONCLUSIONS

It is clear from this chapter that a proper treatment of electrostatic interactions is a complex problem in molecular modeling. Research in this area has resulted in a wealth of information on simulation artifacts arising from methodological difficulties. Much is known about "what we should not do" when designing computer experiments involving DNA and counterions. In comparison, we have very little information on "what we should do" to treat the electrostatics in a realistic manner for such systems. Emerging techniques like Ewald summation have the potential to enhance significantly our ability to produce a more accurate model for DNA–counterion systems, but results from MD simulations on DNA obtained by means of Ewald summation are just beginning to appear in the literature. Based on the available information, it is premature to assess the accuracy of the models resulting from these simulations. Comparative studies of DNA with explicit counterions and water molecules under various modeling conditions and force fields will shed more light on the impact of the new force field parameters and advanced methodologies on the structure of DNA.

ACKNOWLEDGMENTS

We thank Kevin J. McConnell and Dr. Nimala Ramadas for providing results from their MD simulation studies. Dr. Shirley Louise-May provided helpful comments on the manuscript. We thank Thomas A. Darden for his critical comments on the section on Ewald summation.

REFERENCES

1. H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S. H. Hsieh, and A. R. Srinivasan, *Biophys. J.*, 63, 751 (1992). The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids. See Web site at Rutgers University, http://ndbserver.rutgers.edu/.
2. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, 112, 535 (1977). The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures.
3. D. L. Beveridge, S. Swaminathan, G. Ravishanker, J. M. Withka, J. Srinivasan, C. Prevost,

S. Louise-May, D. R. Langley, F. M. DiCapua, and P. H. Bolton, in *Water and Biological Macromolecules*, E. Westhof, Ed., CRC Press, Boca Raton, FL, 1993, pp. 165–225. Molecular Dynamics Simulations on the Hydration, Structure and Motions of DNA Oligomers.

4. W. F. van Gunsteren, *Curr. Opin. Struct. Biol.*, 3, 277 (1993). Molecular Dynamics Studies of Proteins.
5. P. E. Smith and W. F. van Gunsteren, in *Computer Simulation of Biomolecular Systems*, W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, Eds., ESCOM, Leiden, 1993, pp. 182–211. Methods for the Evaluation of Long-Range Electrostatic Forces in Computer Simulations of Molecular Systems.
6. H. J. C. Berendsen, in *Computer Simulation of Biomolecular Systems*, W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, Eds., ESCOM, Leiden, 1993, pp. 161–181. Electrostatic Interactions.
7. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
8. R. E. Dickerson, in *Structure & Methods*, Vol. 3: *DNA & RNA*, R. H. Sarma and M. H. Sarma, Eds., Adenine Press, New York, 1990, pp. 1–38. What Do We Really Know About B-DNA?
9. R. E. Dickerson, *Methods Enzymol.*, 211, 67 (1991). DNA Structure from A to Z.
10. S. Arnott and D. W. L. Hukins, *J. Mol. Biol.*, 81, 93 (1973). Refinement of the Structure of B-DNA and Implications for the Analysis of X-Ray Diffraction Data from Fibres of Biopolymers.
11. S. Arnott, P. J. Campbell Smith, and R. Chandrasekaran, in *CRC Handbook of Biochemistry and Molecular Biology*, G. Fasman, Ed., CRC Press, Cleveland, 1976, Third edition, Nucleic Acids–Vol.II, pp. 411–422. Atomic Coordinates and Molecular Conformations for DNA–DNA, RNA–RNA, and DNA–RNA Helices.
12. S. Arnott, R. Chandrasekaran, D. L. Birdsall, A. G. W. Leslie, and R. L. Ratliffe, *Nature*, 283, 743 (1980). Left-Handed DNA Helices.
13. W. Fuller and A. Mahendrasingam, in *Nucleic Acid Structure*, S. Neidle and E. Westhof, Eds., Macmillan Press, London, 1987, pp. 101–131. X-ray Fibre Diffraction Studies of DNA: Recent Results and Future Possibilities.
14. G. S. Manning, *Q. Rev. Biophys.*, 11, 179 (1978). The Molecular Theory of Polyelectrolyte Solutions with Applications to the Electrostatic Properties of Polynucleotides.
15. M. L. Bleam, C. F. Anderson, and M. T. Record Jr., *Proc. Natl. Acad. Sci. USA*, 77, 3085 (1980). Relative Binding Affinities of Monovalent Cations for Double-Stranded DNA Studied by Sodium-23 NMR.
16. B. H. Zimm and M. Le Bret, *J. Biomol. Struct. Dyn.*, 1, 461 (1983). Counterion Condensation and System Dimensionality.
17. M. R. Reddy, P. J. Rossky, and C. S. Murthy, *J. Phys. Chem.*, 91, 4923 (1987). Counterion Spin Relaxation in DNA Solutions: A Stochastic Dynamics Simulation Study.
18. B. Jayaram, S. Swaminathan, D. L. Beveridge, K. Sharp, and B. Honig, *Macromolecules*, 23, 3156 (1990). Monte Carlo Simulation Studies on the Structure of the Counterion Atmosphere of B-DNA. Variations on the Primitive Dielectric Model.
19. D. J. Patel and L. Shapiro, *Annu. Rev. Biophys. Biophys. Chem.*, 16, 423 (1987). Nuclear Magnetic Resonance and Distance Geometry Studies of DNA Structures in Solution.
20. J. M. Van de Ven and C. W. Hilbers, *Eur. J. Biochem.*, 178, 1 (1988). Nucleic Acids and Nuclear Magnetic Resonance.
21. D. E. Wemmer, *Curr. Opin. Struct. Biol.*, 1, 452 (1991). The Applicability of NMR Methods to the Solution Structure of Nucleic Acids.
22. N. B. Ulyanov and T. L. James, *Appl. Magn. Resonance*, 7, 21 (1994). Statistical Analysis of DNA Duplex Structures in Solution Derived by High Resolution NMR.
23. W. F. van Gunsteren and H. J. C. Berendsen, *Angew. Chem., Int. Ed. Engl.*, 29, 992 (1990). Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry.

24. G. Ravishanker and D. L. Beveridge, Molecular Dynamics Toolchest Version 2.0: Analysis and Graphical Display of Computer Simulation Results on Proteins and Nucleic Acids, Wesleyan University, Middletown, CT, 1995.
25. J. D. Watson and F. H. C. Crick, *Nature*, 171, 737 (1953). A Structure for Deoxyribonucleic Acid.
26. A. G. W. Leslie, S. Arnott, R. Chandrasekaran, and R. L. Ratliff, *J. Mol. Biol.*, 143, 49 (1980). Polymorphism of DNA Double Helices.
27. F. H. Allen, J. E. Davies, J. J. Galloy, O. Johnson, O. Kennard, C. F. Macrae, E. M. Mitchell, G. F. Mitchell, J. M. Smith, and D. G. Watson, *J. Chem. Inf. Comput. Sci.*, 31, 187 (1991). The Development of Versions 3 and 4 of the Cambridge Structural Database System.
28. A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, and A. Rich, *Nature*, 283, 743 (1979). Molecular Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution.
29. R. E. Dickerson and H. R. Drew, *J. Mol. Biol.*, 149, 761 (1981). Structure of a B DNA Dodecamer. II. Influence of Base Sequence on Helix Structure.
30. R. E. Dickerson, M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, O. Kennard, E. von Kitzing, R. Lavery, H. C. M. Nelson, W. K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. M. Soumpasis, C. S. Tung, A. H. J. Wang, and V. B. Zhurkin, *EMBO J.*, 8, 1 (1989). Definitions and Nomenclature of Nucleic Acid Structure Parameters.
31. M. S. Babcock, E. P. D. Pendault, and W. K. Olson, *J. Mol. Biol.*, 237, 125 (1994). Nucleic Acid Structure Analysis. Mathematics for Local Cartesian and Helical Structure Parameters That Are Truly Comparable Between Structures.
32. R. E. Dickerson, Newhelix, Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA.
33. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.*, 6, 655 (1989). Defining the Structure of Irregular Nucleic Acids: Conventions and Principles.
34. D. Bhattacharyya and M. Bansal, *J. Biomol. Struct. Dyn.*, 10, 213 (1992). Groove Width and Depth of B-DNA Structures Depend on Local Variation in Slide.
35. N. Boutonnet, X. Hui, and K. Zakrzewska, *Biopolymers*, 33, 479 (1993). Looking into the Grooves of DNA.
36. E. Stofer and R. Lavery, *Biopolymers*, 34, 337 (1994). Measuring the Geometry of DNA Grooves.
37. R. Lavery, in *Structure and Expression, Vol. 3, DNA Bending and Curvature*, W. K. Olson, M. H. Sarma, R. H. Sarma, and M. Sundaralingam, Eds., Adenine Press, Schenectady, New York, 1988, pp. 191–211. Junction and Bends in Nucleic Acids: A New Theoretical Modelling Approach.
38. M. A. Young, G. Ravishanker, D. L. Beveridge, and H. M. Berman, *Biophys. J.*, 68, 2452 (1995). Analysis of Local Helix Bending in Crystal Structures of DNA Oligonucleotides and DNA–Protein Complexes.
39. L. Nilsson, G. M. Clore, A. M. Gronenborn, A. T. Brunger, and M. Karplus, *J. Mol. Biol.*, 188, 455 (1986). Structure Refinement of Oligonucleotides by Molecular Dynamics with Nuclear Overhauser Effect Interproton Distance Restraints: Application to 5′ d(CGTACG).
40. P. Yip and D. A. Case, *J. Magn. Res.*, 83, 643 (1989). A New Method for Refinement of Macromolecular Structures Based on Nuclear Overhauser Spectra.
41. A. E. Torda and W. F. van Gunsteren, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, Vol. 3, pp. 143–172. Molecular Modeling Using Nuclear Magnetic Resonance Data.
42. G. S. Manning, *J. Chem. Phys.*, 51, 924 (1969). Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions. I. Colligative Properties.
43. G. S. Manning, *Annu. Rev. Phys. Chem.*, 23, 117 (1972). Polyelectrolytes.
44. C. F. Anderson, M. T. Record Jr., and P. A. Hart, *Biophys. Chem.*, 7, 301 (1978). Sodium-23 NMR Studies of Cation–DNA Interactions.

45. M. L. Bleam, C. F. Anderson, and M. T. Record Jr., *Biochemistry*, 22, 5418 (1983). Sodium-23 Nuclear Magnetic Resonance Studies of Cation–DNA Interactions.

46. S. Padmanabhan, B. Richey, C. F. Anderson, and M. T. Record Jr., *Biochemistry*, 27, 4367 (1988). Interaction of an *N*-Methylated Polyamine Analogue, Hexamethonium (2+), with NaDNA: Quantitative N-14 and Na-23 NMR Relaxation Rate Studies of the Cation–Exchange Process.

47. G. S. Manning, K. K. Ebralidse, A. D. Mirzabekov, and A. Rich, *J. Biomol. Struct. Dyn.*, 6, 877 (1989). An Estimate of the Extent of Folding of Nucleosomal DNA by Laterally Asymmetric Neutralization of the Phosphate Groups.

48. M. O. Fenley, G. S. Manning, and W. K. Olson, *Biopolymers*, 30, 1191 (1990). Approach to the Limit of Counterion Condensation.

49. T. G. Dewey, *Biopolymers*, 29, 1793 (1990). A Ligand Binding Model of Counterion Condensation to Finite Length Polyelectrolytes.

50. B. Jayaram and D. L. Beveridge, *J. Phys. Chem.*, 94, 4666 (1990). Free Energy of an Arbitrary Charge Distribution Imbedded in Coaxial Cylindrical Dielectric Continua: Application to Conformational Preferences of DNA in Aqueous Solution.

51. A. Pullman, B. Pullman, and R. Lavery, *J. Mol. Struct.*, 93, 85 (1983). Molecular Electrostatic Potential Versus Field. Significance for DNA and Its Constituents.

52. A. E. Garcia and D. M. Soumpasis, *Proc. Natl. Acad. Sci. USA*, 86, 3160 (1989). Harmonic Vibrations and Thermodynamic Stability of a DNA Oligomer in Monovalent Salt Solution.

53. F. Hirata and R. M. Levy, *J. Phys. Chem.*, 93, 479 (1989). Salt-Induced Conformational Changes in DNA: Analysis Using the Polymer RISM Theory.

54. R. Bacquet and P. J. Rossky, *J. Phys. Chem.*, 88, 2660 (1984). Ionic Atmosphere of Rodlike Polyelectrolytes. A Hypernetted Chain Study.

55. E. Westhof and D. L. Beveridge, in *Water Science Reviews*, F. Franks, Ed., Cambridge University Press, Cambridge, 1990, pp. 24–123. Hydration of Nucleic Acids.

56. R. J. Bacquet and P. J. Rossky, *J. Phys. Chem.*, 92, 3604 (1988). Ionic Distributions and Competitive Association in DNA/Mixed Salt Solutions.

57. W. H. Braunlin and L. Nordenskiöld, *Eur. J. Biochem.*, 142, 133 (1984). Potassium-39 NMR Study of Potassium Binding to Double Helical DNA.

58. W. H. Braunlin, C. F. Anderson, and M. T. Record Jr., *Biochemistry*, 26, 7724 (1987). Competitive Interactions of $[Co(NH_3)_6]^{3+}$ and Na^+ with Helical B-DNA Probed by Cobalt-59 and Sodium-23 NMR.

59. W. H. Braunlin and Q. Xu, *Biopolymers*, 32, 1703 (1992). Hexaamminecobalt(III) Binding Environments on Double-Helical DNA.

60. W. H. Braunlin, T. Drakenberg, and L. Nordenskiöld, *J. Biomol. Struct. Dyn.*, 10, 133 (1992). Ca^{2+} Binding Environments on Natural and Synthetic Polymeric DNA's.

61. S. D. Kennedy and R. G. Bryant, *Biophys. J.*, 50, 669 (1986). Manganese–Deoxyribonucleic Acid Binding Modes.

62. A. P. Williams, C. E. Longfellow, S. M. Freier, R. Kierzek, and D. H. Turner, *Biochemistry*, 28, 4283 (1989). Laser Temperature-Jump, Spectroscopic, and Thermodynamic Study of Salt Effects on Duplex Formation by dGCATGC.

63. D. Rentzeperis, D. W. Kupke, and L. A. Marky, *Biopolymers*, 32, 1065 (1992). Differential Hydration of Homopurine Sequences Relative to Alternating Purine/Pyrimidine Sequences.

64. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, 4, 187 (1983). CHARMM: A Program for Macromolecular Energy, Minimization, Dynamics Calculations.

65. P. S. Subramanian, G. Ravishanker, and D. L. Beveridge, *Proc. Natl. Acad. Sci. USA*, 85, 1836 (1988). Theoretical Considerations on the "Spine of Hydration" in the Minor Groove of d(CGCGAATTCGCG)·d(GCGCTTAAGCGC): Monte Carlo Computer Simulation.

66. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987. J. M. Haile, *Molecular Dynamics Simulation, Elementary Methods*, Wiley, New York, 1992.

67. T. P. Lybrand, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1990, Vol. 1, pp. 295–320. Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods. T. P. Straatsma, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1996, Vol. 9, pp. 81–127. Free Energy by Molecular Simulation.

68. P. Auffinger, S. Louise-May, and E. Westhof, *J. Chem. Soc.*, 117, 6720 (1995). Multiple Molecular Dynamics Simulations of the Anticodon Loop of $tRNA^{Asp}$ in Aqueous Solution with Counterions.

69. D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, D. M. Ferguson, G. L. Seibel, U. C. Singh, P. Weiner, and P. Kollman, *AMBER 4.1*, 1995, University of California at San Francisco, CA.

70. A. Rahman and F. H. Stillinger, *J. Chem. Phys.*, 55, 336 (1971). Molecular Dynamics of Liquid Water.

71. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.*, 79, 926 (1983). Comparison of Simple Potential Functions for Simulating Liquid Water.

72. W. F. van Gunsteren and H. J. C. Berendsen, Groningen Molecular Simulation (GROMOS) 87, Groningen, The Netherlands, 1987. See also Ref. 112 re SPC water model.

73. G. Ravishanker, S. Swaminathan, and D. L. Beveridge, WesDyn 2.0: Molecular Dynamics and Monte Carlo Stimulation for Proteins and DNA, Wesleyan University, Middletown, CT, 1994.

74. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am Chem. Soc.*, 117, 5179 (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.

75. J. Pranata, S. G. Wierschke, and W. L. Jorgensen, *J. Am. Chem. Soc.*, 113, 2810 (1991). OPLS Potential Functions for Nucleotide Bases. Relative Association Constants of Hydrogen-Bonded Base Pairs in Chloroform.

76. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 110, 1657 (1988). The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin.

77. R. J. Loncharich and B. R. Brooks, *Proteins: Struct., Funct., Genet.*, 6, 32 (1989). The Effects of Truncating Long-Range Forces on Protein Dynamics.

78. H. Schreiber and O. Steinhauser, *Biochemistry*, 31, 5856 (1992). Cutoff Size Does Strongly Influence Molecular Dynamics Results on Solvated Polypeptides.

79. P. Auffinger and D. L. Beveridge, *Chem. Phys. Lett.*, 234, 413 (1995). A Simple Test for Evaluating the Truncation Effects in Simulation of Systems Involving Charged Groups.

80. D. H. Kitson, F. Avbelj, J. Moult, D. T. Nguyen, J. E. Mertz, D. Hadzi, and A. T. Hagler, *Proc. Natl. Acad. Sci. USA*, 90, 8920 (1993). On Achieving Better than 1 Å Accuracy in a Simulation of a Large Protein: *Streptomyces griseus* Protease A.

81. N. Ramadas, G. Ravishanker, and D. L. Beveridge, unpublished results, 1990–1993. Critical evaluation of methodological issues in DNA simulations—A case study using aqueous solution of d(CGCGAATTCGCG) with and without explicit counterions.

82. K. J. McConnell, R. Nirmala, M. A. Young, G. Ravishanker, and D. L. Beveridge, *J. Am. Chem. Soc.*, 116, 4461 (1994). A Nanosecond Molecular Dynamics Trajectory for a B DNA Double Helix: Evidence for Substates.

83. P. Ewald, *Ann. Phys*, 64, 253 (1921). Die berechnung optischer und elektrostatischer Gitterpotentiale.

84. S. W. de Leeuw, J. W. Perram, and E. R. Smith, *Proc. R. Soc. London*, A373, 27 (1980). Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants.

85. T. A. Darden, D. M. York, and L. G. Pedersen, *J. Chem. Phys.*, 98, 10089 (1993). Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems.

86. U. Essmann, L. Perera, M. L. Berkowitz, T. A. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.*, 103, 8577 (1995). A Smooth Particle Mesh Ewald Method.

87. T. E. Cheatham III, J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman, *J. Am. Chem. Soc.*, 117, 4193 (1995). Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins.

88. M. A. Young, G. Ravishanker, and D. L. Beveridge, *Biophys. J.*, submitted (1997). Nanosecond Molecular Dynamics Trajectories for a B-DNA Oligonucleotide Based on the AMBER 4.1 Force Field Including Water and Counterions.

89. S. Swaminanthan, G. Ravishanker, and D. L. Beveridge, *J. Am. Chem. Soc.*, 111, 5027 (1991). Molecular Dynamics of B-DNA Including Water and Counterions: A 140 ps Trajectory for d(CGCGAATTCGCG) Based on the GROMOS Force Field.

90. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.*, 6, 63 (1988). The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic Acids.

91. G. Ravishanker, S. Swaminathan, D. L. Beveridge, R. Lavery, and H. Sklenar, *J. Biomol. Struct. Dyn.*, 6, 669 (1989). Conformational and Helicoidal Analysis of 30 ps of Molecular Dynamics on the d(CGCGAATTCGCG) Double Helix: "Curves, Dials and Windows."

92. M. A. L. Eriksson and A. Laaksonen, *Biopolymers*, 32, 1035 (1992). A Molecular Dynamics Study of Conformational Changes and Hydration of Left-Handed $d(CGCGCGCGCGCG)_2$ in a Nonsalt Solution.

93. J. Srinivasan, J. M. Withka, and D. L. Beveridge, *Biophys. J.*, 58, 533 (1990). Molecular Dynamics of an In Vacuo Model of Duplex d(CGCGAATTCGCG) in the B-Form Based on the AMBER 3.0 Force Field.

94. V. Fritsch, G. Ravishanker, D. L. Beveridge, and E. Westhof, *Biopolymers*, 33, 1537 (1993). Molecular Dynamics Simulations of poly(dA)−(dT): Comparisons Between Implicit and Explicit Solvent Representations.

95. S. Falsafi and N. O. Reich, *Biopolymers*, 33, 459 (1993). Molecular Dynamic Simulations of B-DNA: An Analysis of the Role of Initial Molecular Configuration, Randomly Assigned Velocity Distribution, Long Integration Times, and Nonconstrained Termini.

96. C. Prevost, S. Louise-May, G. Ravishanker, R. Lavery, and D. L. Beveridge, *Biopolymers*, 33, 335 (1993). Persistence of the Static and Dynamical Helix Deformations of DNA Oligonucleotides: Application to the Crystal Structure and Molecular Dynamics Simulation of $d(CGCGAATTCGCG)_2$.

97. T. R. Forester and I. R. McDonald, *Mol. Phys.*, 72, 643 (1991). Molecular Dynamics Studies of the Behavior of Water Molecules and Small Ions in Concentrated Solutions of Polymeric B-DNA.

98. V. P. Chuprina, U. Heinemann, A. A. Nurislamov, P. Zielenkiewicz, and R. E. Dickerson, *Proc. Natl. Acad. Sci. USA*, 88, 593 (1991). Molecular Dynamics Simulation of the Hydration Shell of a B-DNA Decamer Reveals Two Main Types of Minor-Groove Hydration Depending on Groove Width.

99. P. S. Subramanian and D. L. Beveridge, *J. Biomol. Struct. Dyn.*, 6, 1093 (1989). A Theoretical Study of the Aqueous Hydration of Canonical B d(CGCGAATTCGCG): Monte Carlo Simulation and Comparison with Crystallographic Ordered Water Sites.

100. P. K. Mehrota and D. L. Beveridge, *J. Am. Chem. Soc.*, 102, 4287 (1980). Structural Analysis of Molecular Solutions Based on Quasi-Component Distribution Functions. Application to $[H_2CO]_{aq}$ at 25 °C.

101. M. Le Bret and B. H. Zimm, *Biopolymers*, 23, 271 (1984). Monte Carlo Determination of the Distribution of Ions About a Cylindrical Polyelectrolyte.

102. C. S. Murthy, R. J. Bacquet, and P. J. Rossky, *J. Phys. Chem.*, 89, 701 (1985). Ionic Distributions Near Polyelectrolytes: A Comparison of Theoretical Approaches.

103. P. Mills, C. F. Anderson, and M. T. Record Jr., *J. Phys. Chem.*, 89, 3984 (1985). Monte Carlo Studies of Counterion–DNA Interactions. Comparison of the Radial Distribution of Counterions with Predictions of Other Polyelectrolyte Theories.

104. P. Mills, M. D. Paulsen, C. F. Anderson, and M. T. Record Jr., *Chem. Phys. Lett.*, 129, 155 (1986). Monte Carlo Simulations of Counterion Accumulation Near Helical DNA.

105. M. D. Paulsen, C. F. Anderson, and M. T. Record Jr., *Biopolymers*, 27, 1249 (1988). Counterion Exchange Reactions on DNA: Monte Carlo and Poisson–Boltzmann Analysis.

106. J. Conrad, M. Troll, and B. H. Zimm, *Biopolymers*, 27, 1711 (1988). Ions Around DNA: Monte Carlo Estimates of Distribution with Improved Electrostatic Potentials.

107. H. R. Drew and R. E. Dickerson, *J. Mol. Biol.*, 151, 535 (1981). Structure of a B-DNA Dodecamer. 3. Geometry of Hydration.

108. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.*, 106, 765 (1984). A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins.

109. B. E. Hingerty, R. H. Ritchie, T. L. Ferrel, and J. E. Turner, *Biopolymers*, 24, 427 (1985). Dielectric Effects in Biopolymers: The Theory of Ionic Saturation Revisited.

110. J. Ramstein and R. Lavery, *Proc. Natl. Acad. Sci. USA*, 85, 7231 (1988). Energetic Coupling Between DNA Bending and Base Pair Opening.

111. H. L. Gordon and S. Goldman, *J. Phys. Chem.*, 96, 1921 (1992). Simulations of the Counterion and Solvent Distribution Functions Around Two Simple Models of Polyelectrolyte.

112. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces. Interaction Models for Water in Relation to Protein Hydration*, B. Pullman, Ed., Reidel, Dordrecht, 1981, pp. 331–342. SPC Water Model.

113. P. A. Mills, A. Rashid, and T. L. James, *Biopolymers*, 32, 1491 (1992). Monte Carlo Calculations of the Ion Distributions Surrounding the Oligonucleotide $d(ATATATATAT)_2$ in the B, A, and Wrinkled D Conformations.

114. L. Pauling, The Nature of the Chemical Bond, Third edition, Cornell University Press, Ithaca, NY, 1960.

115. V. Vlachy and A. D. J. Haymet, *J. Chem. Phys.*, 84, 5874 (1986). A Grand Canonical Monte Carlo Simulation Study of Polyelectrolyte Soutions.

116. B. Jayaram and D. L. Beveridge, *J. Phys. Chem.*, 95, 2506 (1991). Grand Canonical Monte-Carlo Simulations on Aqueous Solutions of NaCl and NaDNA: Excess Chemical Potentials and Sources of Nonideality in Electrolyte and Polyelectrolyte Solutions.

117. D. M. York, W. Yang, H. Lee, T. A. Darden, and L. G. Pedersen, *J. Am. Chem. Soc.*, 117, 5001 (1995). Toward the Accurate Modeling of DNA: The Importance of Long-Range Electrostatics.

118. S. Weerasinghe, P. E. Smith, V. Mohan, Y.-K. Cheng, and B. M. Petitt, *J. Am. Chem. Soc.*, 117, 2147 (1995). Nanosecond Dynamics and Structure of a Model DNA Triplex Helix in Saltwater Solution.

119. S. Vijayakumar and D. L. Beveridge, unpublished data, 1995. MD simulations of lambda repressor protein–DNA complex.

120. D. R. Langley, T. W. Doyle, and D. L. Beveridge, *J. Am. Chem. Soc.*, 113, 4395 (1991). The Dynemicin–DNA Intercalation Complex. A Model Based on DNA Affinity Cleavage and Molecular Dynamics Simulation.

121. D. L. Beveridge, K. J. McConnell, R. Nirmala, M. A. Young, S. Vijayakumar, and G. Ravishanker, in *Structure and Reactivity in Aqueous Solution: Characterization of Chemical and Biological Systems*, C. J. Cramer and D. G. Truhlar, Eds., ACS Symposium Series 568, American Chemical Society, Washington, D. C., 1994, pp. 381–394. Molecular Dynamics Simulations of DNA and Protein–DNA Complexes Including Solvent.

122. D. L. Beveridge and G. Ravishanker, *Curr. Opin. Struct. Biol.*, 4, 246 (1994). Molecular Dynamics Studies of DNA.

123. U. C. Singh, S. J. Weiner, and P. A. Kollman, *Proc. Natl. Acad. Sci. USA*, 82, 755 (1985). Molecular Dynamics Simulations of d(CGCGA)•d(TCGCG) With and Without "Hydrated" Counterions.

124. G. L. Seibel, U. C. Singh, and P. A. Kollman, *Proc. Natl. Acad. Sci. USA*, 82, 6537 (1985). A Molecular Dynamics Simulation of Double-Helical B-DNA Including Counterions and Water.

125. W. F. van Gunsteren, H. J. C. Berendsen, R. G. Geurtsen, and H. R. J. Zwinderman, *Ann. N. Y. Acad. Sci.*, 482, 287 (1986). A Molecular Dynamics Computer Simulation of an Eight-Base-Pair DNA Fragment in Aqueous Solution: Comparison with Experimental Two-Dimensional NMR Data.

126. W. F. van Gunsteren and H. J. C. Berendsen, GROMOS86: Groningen Molecular Simulation System, University of Groningen, Groningen, The Netherlands, 1986.

127. K. N. Swamy and E. Clementi, *Biopolymers*, 26, 1901 (1987). Hydration Structure and Dynamics of B- and Z-DNA in the Presence of Counterions Via Molecular Dynamics Simulations.

128. O. Matsuoka, E. Clementi, and M. Yoshimine, *J. Chem. Phys.*, 64, 1351 (1976). CI Study of the Water Dimer Potential Surface.

129. A. Laaksonen, L. G. Nilsson, B. Joensson, and O. Teleman, *Chem. Phys.*, 129, 175 (1989). Molecular Dynamics Simulation of Double Helix Z-DNA in Solution.

130. E. Clementi and G. Corongiu, *J. Biol. Phys.*, 11, 33 (1983). Structure of Aggregates of Water and Lithium, Sodium or Potassium Counterions with Nucleic Acid in Solution.

131. S. N. Rao and P. A. Kollman, *Biopolymers*, 29, 517 (1990). Simulations of the B-DNA Molecular Dynamics of d(CGCGAATTCGCG) and d(CGCGCGCGCGCG): An Analysis of the Role of Initial Geometry and a Comparison of United and All-Atom Models.

132. T. J. Zielinski and M. Shibata, *Biopolymers*, 29, 1027 (1990). A Molecular Dynamics Simulation of the $(dG)_6$·$(dC)_6$ Minihelix Including Counterions and Water.

133. S. Louise-May and D. L. Beveridge, unpublished work, 1995. Molecular dynamics simulations on the A and B form of d(CGCGAATTCGCG) based on the CHARMM force field.

134. D. R. Langley, G. Ravishanker, and D. L. Beveridge, unpublished data, 1993. MD simulations of d(CCAACGTTCG) using the CHARMM force field.

135. M. Shibata, T. J. Zielinski, and R. Rein, *Biopolymers*, 31, 211 (1991). A Molecular Dynamics Study of the Effect of G·T Mispairs of the Conformation of DNA in Solution.

136. J. Chandrasekhar, D. C. Spellmeyer, and W. L. Jorgensen, *J. Am Chem. Soc.*, 106, 903 (1984). Energy Component Analysis for Dilute Aqueous Solutions of Li^+, Na^+, F^-, and Cl^- Ions.

137. D. H. Powell, A. C. Barnes, J. E. Enderby, G. W. Neilson, P. S. Salmon, *Faraday Disc. Chem. Soc.*, 85, 137 (1988). The Hydration Structure Around Chloride Ions in Aqueous Solution.

138. V. Fritsch and E. Westhof, *J. Chim. Phys.*, 88, 2543 (1991). Molecular Dynamics Simulations of DNA Oligomers Under Various Electrostatic Parameters.

139. V. Fritsch and E. Westhof, *J. Am. Chem. Soc.*, 113, 8271 (1991). Three Center Hydrogen Bonds in DNA: Molecular Dynamics of Poly(dA)*Poly(dT).

140. V. Fritsch and E. Westhof, *J. Comput. Chem.*, 12, 147 (1991). Minimization and Molecular Dynamics Studies of Guanosine and Z-DNA Modified by *N*-2-Acetylaminofluorene.

141. R. M. Venable, G. Wildmalm, B. R. Brooks, W. Egan, and R. W. Pastor, *Biopolymers*, 32, 783 (1992). Conformational States of a TT Mismatch from Molecular Dynamics Simulation of Duplex d(CGCGATTCGCG).

142. K. Miaskiewicz, R. Osman, and H. Weinstein, *J. Am. Chem. Soc.*, 115, 1526 (1992). Molecular Dynamics Simulation of the Hydrated d(CGCGAATTCGCG) Dodecamer.

143. A. A. Lipanov and V. P. Chuprina, *Nucleic Acid Res.*, 15, 5833 (1987). The Structure of poly(dA):poly(dT) in a Condensed State and in Solution.