COMPUTATIONAL CHALLENGES IN SIMULATING LARGE DNA OVER LONG TIMES

Simulating DNA's dynamics requires a sophisticated array of algorithms appropriate for DNA's impressive spectrum of spatial and temporal levels. The authors describe computational challenges, solution approaches, and applications that their group has performed in DNA dynamics.

The venerable deoxyribonucleic acid molecule has not always held the spotlight. From Gregor Mendel's first careful work in 1865 to the painstaking evidence Oswald Avery, Alfred Hershey, and Marsha Chase and colleagues obtained from 1944 through 1954, we have experienced dazzling progress in our appreciation of DNA and our ability to read, interpret, and manipulate heredity's master molecule. Just short of a household word, DNA now plays a key role in medical diagnoses through gene markers, bioengineering and nanotechnology constructs, historical analyses, crime and forensics, and family lineage verifications, to name a few.

As we embark on a new millennium, emerging genomic research areas seek to characterize gene products and relate them among species, and expand our interest beyond a single molecule to integrated cellular structures and functions. At least two features are key to achieving these important giant leaps of genome integration in the coming decade. The first is a better ability to compute the 3D structures of biomolecules from the primary sequence (amino acids in proteins and nucleotides in nucleic acids). The second is the development of efficient computational technologies and strategies to analyze sequences, structures, and functions. We rely crucially on such tools to extract knowledge from the wealth of emerging database information on biomolecules.

Although attention has focused on protein structure and folding (how single-stranded polynucleotides fold back on themselves to form complex 3D molecular architectures), analogous problems in DNA and its cousin ribonucleic acid (RNA) are at least as important and perhaps even more challenging. Unlike the relatively compact structure of globular proteins, DNA has many levels of structural hierarchy, from length scales of nanometers for several base pairs to micrometers for several thousand base pairs. Crucial for recognition by proteins, DNA's sequence-dependent behavior on the base-pair level must be analyzed. The study of large-scale DNA folding on the thousand-base-pair level is also of great interest because of its importance to the packaging of the genome into chromosomes and the associated biological regulation processes.

Given these broad goals, studying DNA begs for multidisciplinary collaborations that involve not only chemists and biologists but also mathematicians and other physical scientists. In this article, we describe some of the computational

1521-9615/99/$10.00 © 1999 IEEE

TAMAR SCHLICK
New York University and Howard Hughes Medical Institute
DANIEL A. BEARD, JING HUANG, DANIEL A. STRAHS, AND XIAOLIANG QIAN
New York University

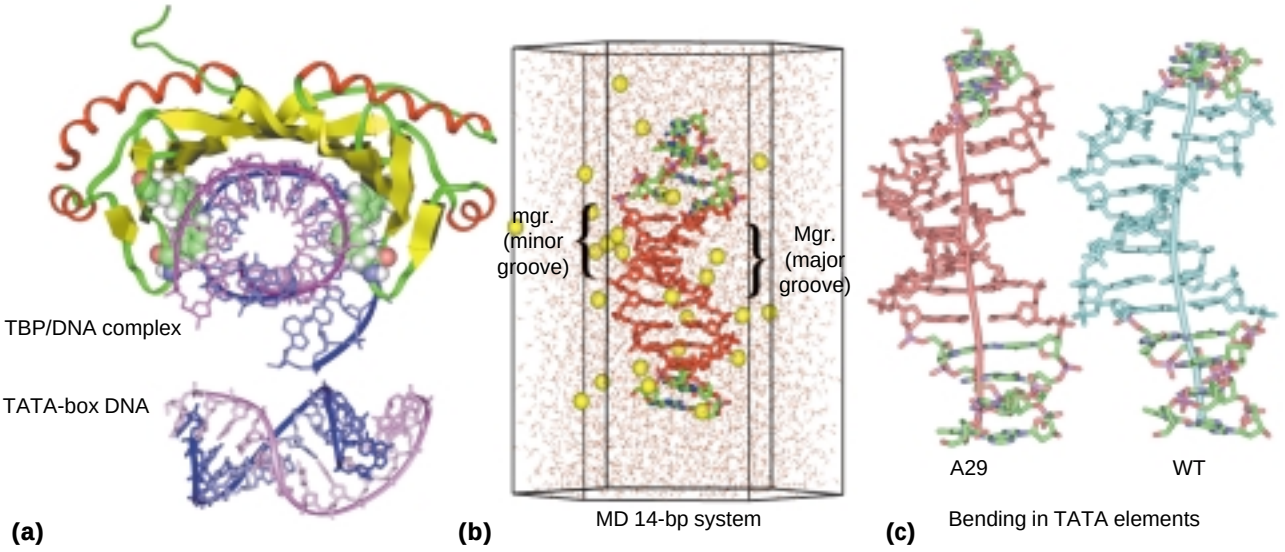

Figure 1. (a) The TATA-box binding protein (TBP), bound to wildtype adenovirus DNA (with central 5' TATAAAAG 3' sequence), whose coordinates are available from the crystal structure. The distorted DNA element from this co-crystal is rotated to highlight its 90° bend. (b) The molecular-dynamics simulation cell (hexagonal prism) of the AdMLP DNA element. The phosphate-neutralizing sodium ions are yellow and the water molecules are faint red-and-white sticks. The major (Mgr.) and minor (mgr.) grooves are also shown. (c) The two computed MD-ensemble averages of the TATA box-containing DNA systems of 14 base pairs, the wildtype sequence (WT), and its single base pair variant 5' TAAAAAAG 3' (A29, an adenine-rich DNA sequence known as an A-tract). The TATA box is indicated in blue (A29) or red (WT), and the global helical axes are illustrated for each system.

challenges of simulating DNA's dynamics, focusing on the large-scale and long-time modeling work in our group. These approaches incorporate chemistry and biology as well as elements of mathematical topology and geometry, elasticity theory, mechanics, and scientific computing.

The DNA molecule and its inherent flexibility

The classic DNA double helix that Francis Crick and James D. Watson described in 1953 is a flexible ladder-like structure of two intertwined polynucleotide chains running in anti-parallel fashion. The nucleotide building block consists of sugar (deoxyribose), phosphate, and base units. One strand runs from the C5′–OH group of the first sugar to the C3′–OH group of the last, while the complementary strand runs from C3′–OH of the first sugar group's partner to the corresponding C5′–OH end of the last base. The ladder rails comprise alternating sugars and phosphates, and each ladder rung is a nitrogenous base pair held together by two or three hydrogen bonds. Adenine (A) often pairs with thymine (T), and guanine (G) frequently pairs with cytosine (C). The spaces formed between the helical backbone and the imaginary cylinder that encloses the DNA are termed *major* and *minor grooves*; they have different dimensions because of the sugar-based linkages' asymmetry with respect to the base-pair plane (see Figure 1).

We use standard atom and dihedral-angle labeling schemes for nucleic acids. The sequence of nitrogenous bases in the 5′ to 3′ strand specifies the DNA's composition; thus, the sequence 5′ TATAAAAG 3′ implies the complementary strand 5′ CTTTTATA 3′. Besides A–T and G–C base pairs, researchers have observed many other hydrogen-bonding patterns for normal and modified bases, especially in RNA molecules. (RNAs have uracil (U) instead of thymine, and ribose instead of deoxyribose.) Other references provide excellent introductions to DNA structure.[1,2]

DNA's 3D structure depends on many factors: base composition, environmental conditions (such as relative humidity and salt concentration), and the presence of other molecules that interact with DNA (such as proteins or drugs). As in proteins, the DNA sequence contains subtle information on local variations that can become collectively pronounced over large spatial scales. Sequence-dependent variations are manifested by rotational and translational deformations from ideal helical orientations (in which the base pairs

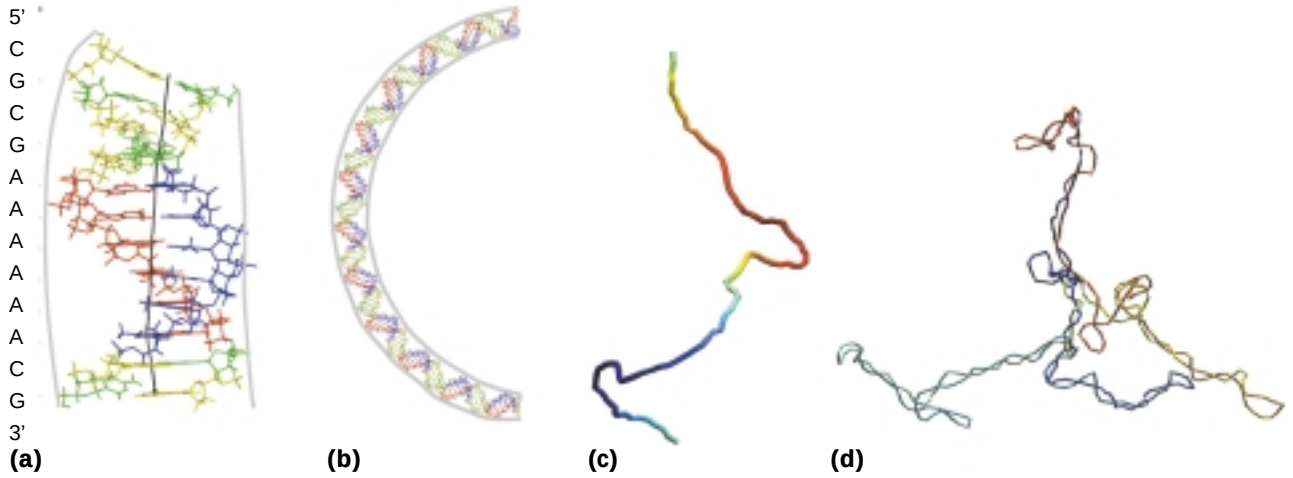

Figure 2. Models of DNA at four different length scales: (a) an A-tract dodecamer with an overall curvature of 11°, (b) a model of 120 base pairs of a phased A-tract sequence, (c) linear DNA of 1.2 kbp, and (d) supercoiled DNA of 12 kbp. Our computed dodecamer by all-atom molecular dynamics served as the model for constructing the 120-base-pair system; the larger linear and supercoiled structures are representative of the thermal equilibrium ensemble, as generated by Brownian dynamics simulations. The curve for the long DNA represents the double helix.

are all perpendicular to the global helical axis). Average roll and tilt rotations—deformations along the long and short base-pair axes, respectively—are generally a few degrees in protein-free DNA but can be more pronounced for DNA bound to proteins (see Figure 1). Twist—the rotation along the global helical axis from one base pair to the next—exhibits a range of values below and above the average of 34° associated with the 10.5 base pair/turn repeat of canonical B-DNA in solution. (B-DNA is the classic, most commonly observed structural form of DNA.) Other translational deformations from the idealized structure identify the base pairs' locations with respect to the global helical axis.[1] Since the two hydrogen-bonded base pairs may themselves deviate from planarity—we also observe a nonzero propeller twist angle, which can be large as 20° or more in certain sequence environments.

As an example of sequence effects, consider the intrinsically curved DNA in Figure 2 that results from adenine-rich sequences in which five or six consecutive adenines are phased with the helical repeat (A-tracts), as Donald Crothers and his colleagues first discovered in the early 1980s.[3] The global helical curve's overall bending is not large on the dodecamer level (11°), but it is pronounced when the sequence pattern, and thus bending propensity, repeats. Figure 2 also shows two other lengths of DNA—linear DNA of 1.2 kbp and supercoiled DNA of 12 kbp (kbp is thousands of base pairs). From the figure, we can view short DNA as a relatively stiff and straight rod, while large DNA resembles flexible polymers undergoing Brownian motion.

Besides the sequence's profound effect on DNA structure, the molecule's architecture as a whole—handedness, helical geometry, and so on—is sensitively affected by the environment. For example, the canonical B-DNA Crick and Watson described was deduced from X-ray diffraction analyses of the sodium salt of DNA fibers at 92% relative humidity. Another right-handed form of DNA—now termed A-DNA—emerged from early fiber diffraction data at the much lower value of 75% relative humidity. This alternative helical geometry is prevalent in double-helical RNA structures. The peculiar left-handed DNA helix, termed Z-DNA for its zigzag design, was discovered in the 1970s in C–G polymers at high salt concentrations. Its biological significance remains uncertain, but evidence suggests that the conversion from B to Z-like DNA acts as a genetic regulator.

Beyond these three canonical helical forms, we now recognize numerous variations in polynucleotide structures—both helical and nonhelical forms—duplexes, triplexes, quadruplexes, as well as parallel DNA, and hybrids of RNA, DNA, and other polymers.[4] Still, B-DNA is thought to be the dominant form under physiological conditions. One reason for its prevalence is that the B-DNA helix can smoothly bend about itself to form a (left-handed) superhelical structure (*plectoneme*, or interwound structure) with minimal changes in the local

structure (see Figure 2d). This property facilitates distant interactions in long DNA, the packaging of long stretches of genomic DNA in the cell—by promoting volume condensation and protein wrapping—and template-directed processes such as replication and transcription that require the DNA to unwind.[5] (The genome content [in total base pairs] varies from organism to organism but roughly increases with the number of different cell types. For example, bacterial genomes have approximately 10^6 to 10^7 base pairs, but mammals contain approximately 10^9 base pairs. Because the eukaryotic nucleus size is approximately 5 µm [also the cell size in prokaryotes], a value much smaller than the length associated with that amount of stretched DNA, five orders of magnitude of DNA condensation must occur.)

DNA's two levels of resolution

Two levels of DNA structure form the central focus of molecular-simulation research: nucleotide (or base pair) level and kilobase pair level. The former involves the study of a dozen or so base pairs, focusing on sequence effects and local interactions between DNA and proteins or other biomolecules. The latter involves long circular or linear DNA, focusing on global structure and folding kinetics related to biological processes such as site-specific recombination.[6,7] High-resolution methods such as nuclear magnetic resonance and crystallography for structure determination guide atomic-level models. Lower-resolution techniques such as gel electrophoresis and electron microscopy provide information on supercoiled DNA.

MD applications

An example in the first application area is intrinsically bent, adenine-rich DNA studied with all-atom molecular dynamics (MD). Simulations produce insights into the controversial relationship between crystallographic and solution data of DNA A-tracts and the forces that stabilize bending. Specifically, research supports preferential bending of A-tracts into the minor groove,[8,9] and consolidates experimental observations concerning the departure of some crystal models from this orientation.[10]

MD simulations of protein-binding DNA sequences that vary by a single base pair from each other have helped interpret experimental data[11,12] regarding the relation between the sequence of the DNA promoter and the biological transcriptional activity of DNA–protein complexes. Specifically, many groups[13–15] have simulated short DNA segments called TATA elements that bind to the TATA box-binding protein (TBP); this binding is a prerequisite for transcription initiation in eukaryotes.[11] Significantly, protein binding imposes a large distortion on DNA. However, the protein succeeds in inducing this enormous deformation because of the DNA's incisive cooperation: evolution has apparently selected the TATA box element 5′ TATAAAAG 3′ found in adenovirus because of its inherent flexibility.[16]

Our recent simulations of 13 single base pair TATA variants[16] have revealed several features of this sequence-dependent deformability:

- the preferred TATA sequence bends flexibly into the DNA's major groove, commensurate with the protein deformation (Figure 1);
- optimal backbone shielding by counterions supports this bending;
- a disordered water–DNA interface further facilitates this motion and thus TBP binding; and
- specific local motions at the TATA ends are associated with high-activity sequences.

Intrinsic curvature

Among the many questions addressed by computational scientists studying DNA are the effects of intrinsic curvature on DNA conformation,[17] DNA site juxtaposition[18] (the close spatial approach of linearly distant regions), and chromatin folding.[19,20] DNA site juxtaposition brings together in space linearly distant DNA segments. Many reactions such as site-specific recombination and transcription depend on such a spatial approach; in some cases, this interaction only occurs if the DNA is supercoiled. Simulations help us understand the reasons for this requirement, the mechanisms involved in juxtaposition, and the dependence of site juxtaposition on the level of DNA superhelicity, salt concentration, site separation, and DNA length (see the "Site juxtaposition kinetics" sidebar).

Modeling chromatin folding involves studying the dynamics of the nucleoprotein complex that compacts the genomic material in eukaryotic cells. Dynamics of this spool-like complex (made of DNA wrapped around histone protein cores) plays a key role in regulating basic cellular processes such as chromosomal condensation and replication. The 11-nanometer nucleosome core particle's crystal structure from 1997[21] was a tour de force of structural biology, but how higher-order forms are organized remains a mystery. In

Site juxtaposition kinetics

Our recent investigations into supercoiled DNA dynamics have focused on understanding the juxtaposition mechanism of linearly distant sites along the DNA contour and how variations in the superhelical density and salt concentration affect the process. Juxtaposition of linearly distant sites, which occurs on the time scale of milliseconds, is required for a variety of processes, including site-specific recombination and certain transcriptional events. However, current experimental techniques cannot probe the kinetics involved in great detail.[1,2] Surprisingly, we find that the site juxtaposition mechanism depends critically on the salt concentration. At low salt, we identify random collision as the dominant mechanism, but at high salt, juxtaposition proceeds by slithering[3] (the random reptational, bidirectional movements of the two opposite segments along the superhelical axis) coupled to branching rearrangements of the DNA supercoil.

Specifically, our simulations show that at low salt concentrations and at low DNA superhelical densities, the DNA structure is more irregular. Such loose supercoiling enhances flexibility—the DNA structure undergoes large global superhelical distortions. Because supercoiling increases the equilibrium probability of juxtaposition two orders of magnitude,[4] at low salt concentrations we observe an increase in site juxtaposition rates with superhelical density commensurate with the increase in juxtaposition probability.

In contrast, at physiological concentrations (relatively high salt), the site juxtaposition rate is determined by the combined effects of slithering, branch creation and deletion, and interbranch collisions, and is not sensitive to the changes in the superhelicity.[5] Here, circular DNAs adopt regular, tightly interwound superhelical structures, usually branched for DNA larger than 3 kbp (see Figure A). In such branched DNA structures, these three processes combine to accelerate the site juxtaposition process.

Theoretical analyses[6] of site juxtaposition, assuming purely reptational slithering, reveal an average collision time that scales with $\mathcal{L}^3$. More realistic motions that involve branch creation and deletion along with slithering result[7] in juxtaposition times that scale approximately as $\mathcal{L}^2$. Our simulations suggest a near-quadratic length dependence of site juxtaposition rates at high salt conditions.[5] Hence, at physiological conditions, the juxtaposition rate is not sensitive to the changes in the equilibrium juxtaposition proba-

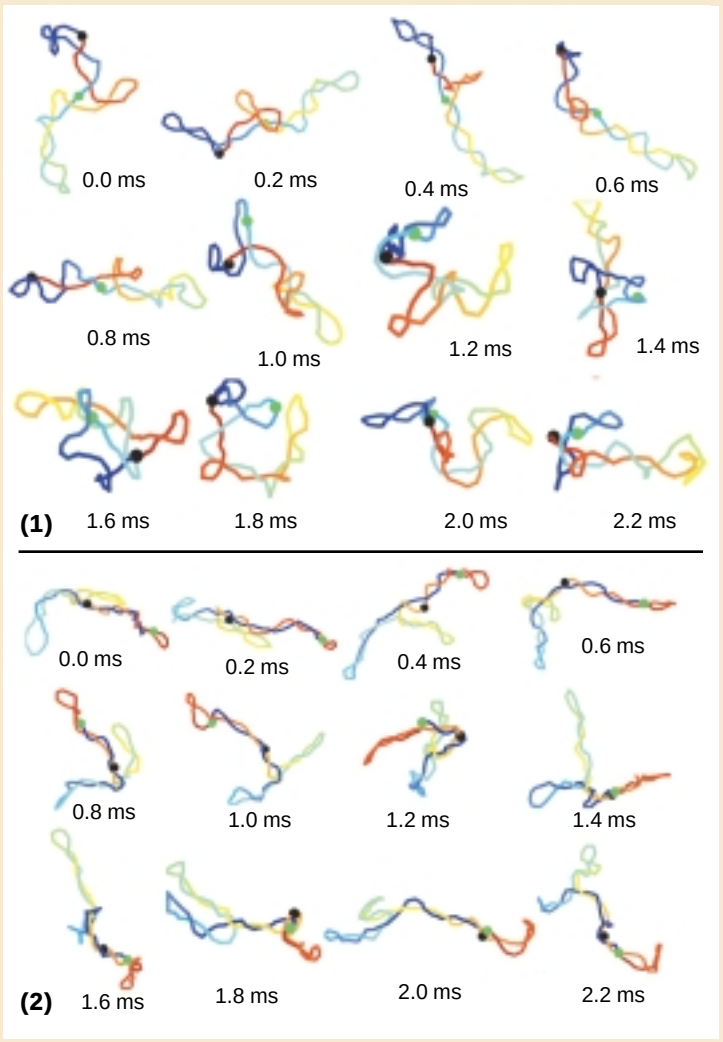

Figure A. Brownian-dynamics snapshots of 3 kbp circular DNA with superhelical density $\sigma = -0.06$ under both (1) low (0.01 M) and (2) high (0.20 M) salt conditions. Discrete 30 base-pair segments model the DNA.

particular, elucidating the details of the transition between the more open and more compact structure will help us better understand transcriptional regulation and DNA packaging (see the "Chromatin folding simulations" sidebar).

Modeling challenges

The different nature of these levels and associated problems requires different computational apparatuses. Namely, the nucleotide level is usually investigated with all-atom molecular mechanics and dynamics protocols,[6] while the kilobase pair level is studied by macroscopic models investigated with Monte Carlo, Brownian, and Langevin dynamics.[22] The all-atom approach faces the challenge of large system sizes in fully

bility, as deduced from Monte Carlo work.

Figure A illustrates the juxtaposition kinetics at these two salt conditions. At the low concentration (series A1), juxtaposition of two sites (indicated by black and green spheres) proceeds through a rearrangement of the global structure. At the higher salt concentration (series A2), the intertwined structure is fairly regular, and juxtaposition proceeds by slithering and branch sliding. In particular, a three-branch structure remains fairly stable at high salt while thermal motions result in more drastic rearrangements for the low-salt case. At high salt, the highlighted beads gradually slither toward one another and remain in close proximity from 1.4 ms to 2.0 ms, while at low salt the juxtaposition event (occurring at 2.0 ms) is short-lived.

References

1. C.N. Parker and S.E. Halford, "Dynamics of Long-Range Interactions on DNA: The Speed of Synapsis during Site-Specific Recombination of Resolvase," *Cell*, Vol. 66, No. 4, Aug. 1991, pp. 781–791.
2. R.B. Sessions et al., "Random Walk Models for DNA Synapsis by Resolvase," *J. Molecular Biology*, Vol. 270, No. 3, July 1997, pp. 413–425.
3. K.R. Benjamin et al., "Contributions of Supercoiling to Tn3 Resolvase and Phage Mu Gin Site-Specific Recombination," *J. Molecular Biology*, Vol. 256, No. 1, Feb. 1996, pp. 50–65.
4. A.V. Vologodskii and N.R. Cozzarelli, "Effect of Supercoiling on the Juxtaposition and Relative Orientation of DNA Sites," *Biophysical J.*, Vol. 70, No. 6, June 1996, pp. 2548–2556.
5. J. Huang, T. Schlick, and A. Vologodskii, "Dynamics of Site Juxtaposition in Supercoiled DNA," to be published in *Proc. Nat'l Academy of Science USA*, 2000; schlick@nyu.edu.
6. J.F. Marko and E.D. Siggia, "Statistical Mechanics of Supercoiled DNA," *Physical Rev. E*, Vol. 52, No. 3, Sept. 1995, pp. 2912–2938.
7. J.F. Marko, "The Internal 'Slithering' Dynamics of Supercoiled DNA," *Physica A*, Vol. 244, 1997, pp. 263–277.

solvated models, sensitivity to force-field and simulation protocol, accurate treatment of long-range electrostatic interactions, and limitation of simulation times and hence configurational sampling range. The macroscopic representation is limited by model approximations, treatment of hydrodynamic forces and ionic effects, and propagation methods. Both levels are thus challenged by fundamental model assumptions and large computational requirements.

Table 1 shows typical setups and computational requirements for these two types of models. Figure 3 shows the percentage of computational work for different program components. In all-atom molecular and Langevin dynamics protocols, the iterative updating procedure for defining coordinates and momenta is relatively simple, even in multiple time step (MTS) methods, and most of the work involves energy and force evaluation at each time step. The most expensive part of this calculation involves the long-range Coulomb potentials and associated forces. Although this task has largely been accelerated with fast adaptive multipole or Ewald-type methods that approach near-linear complexity with size N (typically $\mathcal{O}(N \log N)$), the time step limitation (femtosecond-order time steps) dictates millions of steps to span a relatively short time in a biomolecule's life. MTS methods for both Newtonian and Langevin dynamics combined with efficient implementations on parallel platforms have also helped alleviate this computational burden,[23–26] letting us simulate larger system sizes over longer times. Recent work on alleviating resonance instabilities by the LN algorithm[27,28] has extended time step values to well over 10 fs for the slow forces, with net speedups as indicated in Table 1 and Figure 3. Still, the computational requirements for atomic-level detail remain large. Currently, we can only accomplish longer simulation times for small systems with simplified long-range force treatments and dedicated supercomputing time.[29]

In Brownian-dynamics (BD) simulations of supercoiled DNA, the propagation equations that dictate each set of coordinates are fairly complex when torsional motion and hydrodynamic forces are involved—elaborations on the standard Ermak and McCammon scheme[30] are necessary.[31] Prescribing the motion essentially requires a prediction–correction step because each discrete segment's rotation is coupled to the movement of the associated bead's local coordinate frame. Incorporating hydrodynamics effects entails solving a dense linear system that involves the configuration-dependent hydrodynamic tensor to define the random force at each step. As we discuss later, this task is generally accomplished by a Cholesky factorization, which increases as $\mathcal{O}(N^3)$ with system size. Although the electrostatic forces dominate the computational time for small and moderately sized DNA systems, the work associated with hydrodynamics dominates for large systems (see Figure 3). Here

Chromatin folding simulations

Another interesting application involves modeling chromatin, the nucleoprotein complex that compacts the genomic material in eukaryotic cells. The chromatin fiber is composed of a chain of globular histone protein octamers connected by linker DNA segments. Continuous with the linker DNA is a 150-base-pair left-handed supercoil of DNA that is wrapped around each octamer. The entire repeating unit of a core particle (octamer plus wrapped DNA) and linker DNA is denoted the *nucleosome*. Chromatin condenses into a compact form, which is a critical regulator of transcription and replication.

This system's size demands a biophysical description in the spirit of polymer-level models of DNA.[1–3] The core protein complex, however, is much less regular in terms of shape and charge distribution than simple DNA. To model the electrostatic interactions in this complex, we developed an algorithm for optimizing a discrete *N*-body Debye-Hückel potential to match the electric field predicted by the nonlinear Poisson-Boltzmann equation.[4] The nucleosome

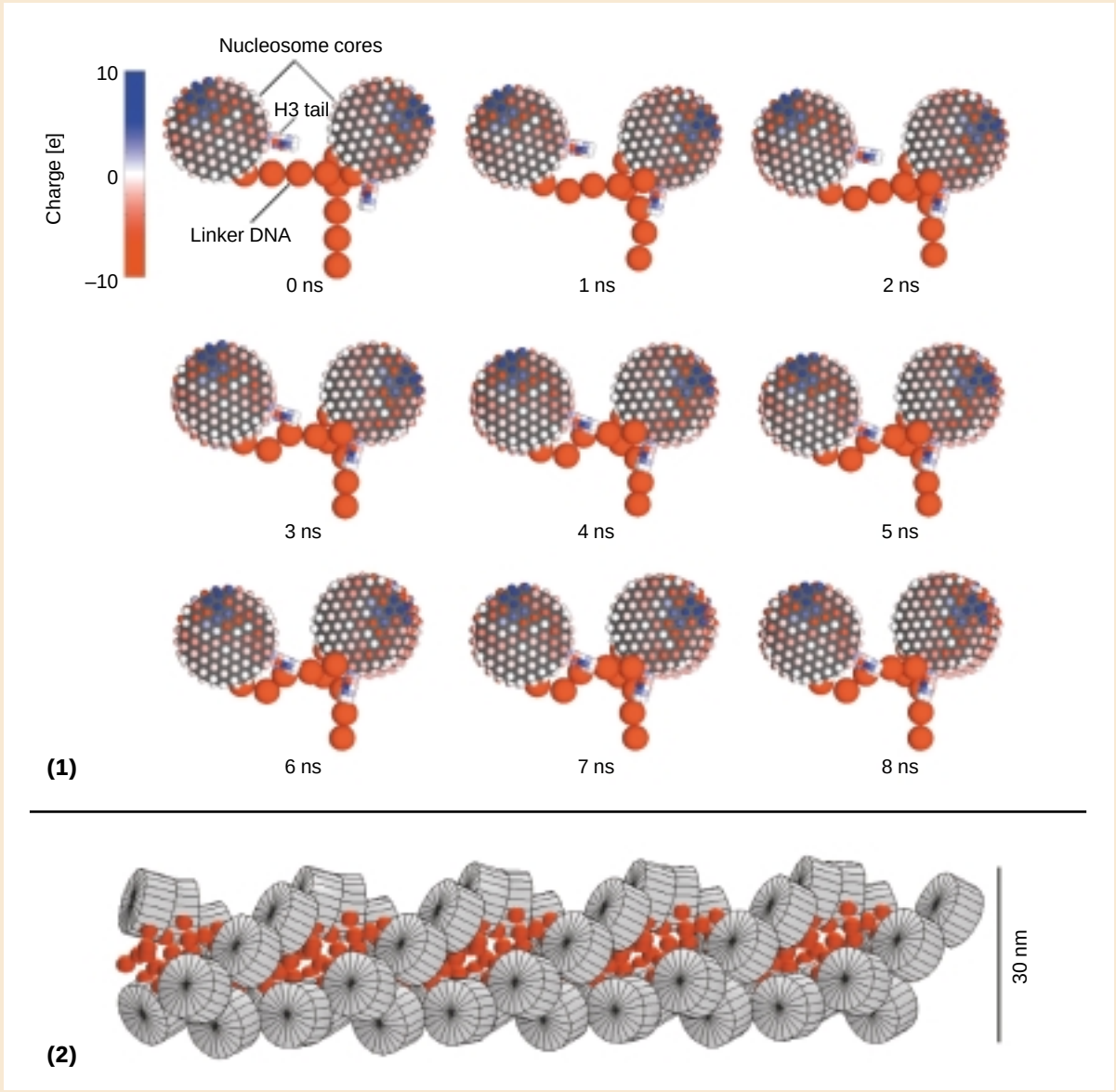

Figure B. Chromatin modeling based on our dinucleosome model of two electrostatically charged core particles connected by an 18-nm linker DNA modeled as an elastic wormlike chain (top left corner). The dinucleosome folding trajectory in part (1) reveals spontaneous folding into a condensed structure in a few nanoseconds. We refined the 30-nm fiber (48 nucleosome units) constructed as a solenoid from the dinucleosome fold motif using Monte Carlo methods to obtain the structure in part (2).

particles in Figure B use the 277-point charge model that we incorporate into a macrolevel polynucleosome model. In this way, a biomolecule's atomic-level details are efficiently integrated into an accurate biophysical description of a system too large to treat on the atomic scale. Energy parameters for the DNA (charge density and elasticity constants) are adopted from studies of DNA supercoiling. We have tested the resulting model parameters against available experimental data, such as translational diffusion constants from chicken erythrocyte polynucleosomes under varying salt concentrations.[5]

In Figure B1 we plot a 4-ns trajectory representing the folding of a two-nucleosome system at monovalent salt concentration of C_s = 0.05 M. The N-terminal H3 tail is positively charged and associates with the negatively charged linker DNA. The linker DNA adopts a bent configuration. Based on the observed fold motif for this system, we can construct larger systems, such as the 48-nucleosome fiber in Figure B2. The predicted fiber is a right-handed solenoid with a diameter of approximately 30 nm, in agreement with experimental observations on chromatin. Our work continues to explore the internal structure of the 30-nm fiber and to interpret folding and unfolding processes associated with acetylation and phosphorylation of the histone proteins.[5]

References

1. W.K. Olson and V.B. Zhurkin, "Modeling DNA Deformations," *Current Opinion in Structural Biology*, Vol. 10, No. 3, June 2000, pp. 286–297.
2. J.A. Martino, V. Katritch, and W.K. Olson, "Influence of Nucleosome Structure on the Three-Dimensional Folding of Idealized Minichromosomes," *Structure with Folding & Design*, Vol. 7, No. 8, Aug. 1999, pp. 1009–1022.
3. L. Ehrlich et al., "A Brownian Dynamics Model for the Chromatin Fiber," *Computer Applications in the Biosciences*, Vol. 13, No. 3, June 1997, pp. 271–279.
4. D. Beard and T. Schlick, "Modeling Salt-Mediated Electrostatics of Macromolecules: The Algorithm DiSCO (Discrete Charge Surface Charge Optimization) and Its Application to the Nucleosome," *Biopolymers*, Vol. 58, 2001; schlick@nyu.edu.
5. D. Beard and T. Schlick, "Computational Modeling Predicts the Structure and Dynamics of the Chromatin Fiber," submitted to *Structure with Folding & Design*, 2000; schlick@nyu.edu.

we show that an alternative algorithm (which Marshall Fixman proposed over a decade ago[32]) dramatically reduces computational times for BD simulations of long DNA. Other recent applications are described elsewhere.[33,34]

The elastic model for long DNA

The elastic-rod approximation has proven valuable for studying superhelical DNA's global features (such as long range and time flexibility). Using ideas from polymer physics, we can characterize long DNA by its contour length $\mathcal{L}$ and a bending rigidity A. We can relate the DNA's mean square displacement $\langle R^2 \rangle$ to the persistence length p_b, which is essentially the length scale on which the polymer directionality is maintained:

$$\langle R^2 \rangle = 2\, p_b\, \mathcal{L}\ . \tag{1}$$

Thus, for lengths $\ll p_b$, we can consider the DNA to be straight, but for lengths $\gg p_b$, a better description is a bent random coil undergoing Brownian motion. This length-dependent flexibility is apparent from Figure 2, which shows DNA on length scales much smaller and much greater than p_b. The persistence length of DNA in vivo is approximately 50 nm, or approximately 150 base pairs at physiological monovalent salt concentrations. The persistence length is also related to the bending force constant A as

$$A = p_b\, k_B\, T \tag{2}$$

where k_B is Boltzmann's constant and T is the temperature. Thus, the floppy polymer writhes through space as a wormlike chain, with the bending rigidity—which tries to keep the DNA straight—balanced by thermal forces—which tend to bend it in all directions.

We can write the elastic-deformation energy as a sum of bending and twisting potentials, with bending and torsional-rigidity constants (A and C) deduced from experimental measurements of DNA bending and twisting.[35] Similar to Equation 2, the torsional rigidity C is related to the twisting persistence length p_{tw} by

$$\mathrm{C} = (p_{tw}\, k_B\, T)\,/2. \tag{3}$$

The bending constant does not have the 1/2 factor because bending involves two axial components of the deformation (roll and tilt) perpendicular to the global helical axis.

Table 1. The complexity of DNA dynamics simulations.

Resolution	System and size (N)	Technique and protocol	Simulation range	CPU performance
All-atom solvated TBP/DNA	37,700 atoms	MD, Leapfrog, Δt = 1 fs	10 ns	787 days/4 processors
All-atom solvated TBP/DNA	37,700 atoms	LD, LN, Δt: 1/2/120 fs	10 ns	121 days/4 processors
Macroscopic DNA, 30 base pairs per bead	Supercoiled DNA, 12 kbp (400 beads)	Second-order BD, hydrodynamics, Δt: 600 ps	10 ms	110 days
Macroscopic DNA, 8 base pairs per bead	Linear DNA, 1.2 kbp (150 beads)	Second-order BD, hydrodynamics, Δt: 600 ps	10 ms	10.9 days
Macroscopic DNA/ protein, 8 base pairs per bead & protein core	48 nucleosomes & linker DNA (240 DNA beads, 48 core beads)	Monte Carlo	1 million steps	60 days

MD stands for molecular dynamics, LD for Langevin dynamics, and BD for Brownian dynamics. All computations are reported on an SGI Origin 2000 with 300-MHz R12000 processors. The LN scheme, named for its origin in a Langevin normal-mode approach, combines force splitting by extrapolation with Langevin dynamics to alleviate severe resonances and allow large outer time steps.[27]

The bending term is proportional to the square of the curvature κ and the twisting energy is proportional to the twist deformation:

$$E = E_B + E_T = \frac{A}{2}\oint \kappa^2(s)\mathrm{d}s + \frac{C}{2}\oint(\omega - \omega_0)^2\mathrm{d}s. \quad (4)$$

In these equations, s denotes arc length, and the integrals are computed over the entire closed DNA curve of length $\mathcal{L}_0$. The DNA's intrinsic twist rate is ω_0 (such as $2\pi/10.5$ radians between successive base pairs). In addition to these bending and twisting deformations, other components account for stretching interactions, electrostatic (screened Coulomb in the form of Debye-Hückel), and hydrodynamic interactions (see the sidebar "A computational model for supercoiled DNA").

BD propagation algorithm and hydrodynamics

To simulate long-time trajectories of DNA motion,[36] researchers commonly use Donald Ermak and J. Andrew McCammon's[30] BD algorithm. The algorithm updates particle positions according to

$$X^{n+1} = X^n + \Delta t \sum_j \frac{\partial D_{ij}}{\partial X_j}\bigg|_{r^n} + \left(\frac{\Delta t}{k_B T}\right)\mathrm{D}(X^n)\cdot f^n + R^n \quad (5)$$

where X^n denotes the collective position vector for the N particles at the nth time step (time $n\Delta t$), f^n is the systematic force (negative gradient of the potential energy), $\mathbf{D}(X^n)$ is the configuration-dependent diffusion tensor, and D_{ij} is the ijth entry of $\mathbf{D}(X^n)$. The allowable time step Δt for BD is typically in the range of 100 picoseconds, orders of magnitude greater than the subfemtosecond time steps used in all-atom MD.

The random-displacement vector R^n is included to mimic thermal interactions with the solvent. It is a Gaussian white noise process related to $\mathbf{D}$ with covariance structure given by

$$\langle R^n\rangle = 0, \langle (R^n)(R^n)^T\rangle = 2\Delta t\ \mathbf{D}(X^n). \quad (6)$$

($\langle (R^n)(R^m)^T\rangle = 0$ for $m \neq n$.) In the BD algorithm, the diffusion tensor $\mathbf{D}$ defines the hydrodynamic interactions among the particles, as well as the correlation structure of the random motions. A reasonable choice for the mathematical form of $\mathbf{D}$ is the Rotne-Prager hydrodynamic tensor,[37] which represents a second-order approximation for two beads diffusing in a Stokes fluid. (The $\sum \partial D_{ij}/\partial X_{ij}$ term in Equation 5 is also zero for this tensor.)

Obtaining this random force R^n in BD algorithms turns out to be the computational bottleneck. We can compute a vector R^n with covariance specified by Equation 6 according to

$$R^n = 2(\Delta t)^{1/2}\mathbf{L}z \quad (7)$$

where z is a vector of uncorrelated random numbers chosen from a Gaussian distribution with zero mean and unit variance (that is, $\langle z^n(z^m)^T\rangle = \delta_{nm}\mathbf{I}$, $\langle z^n\rangle = 0$). The matrix $\mathbf{L}$ comes from a Cholesky factorization of $\mathbf{D}^n$:

$$\mathbf{D} = \mathbf{L}\mathbf{L}^T. \quad (8)$$

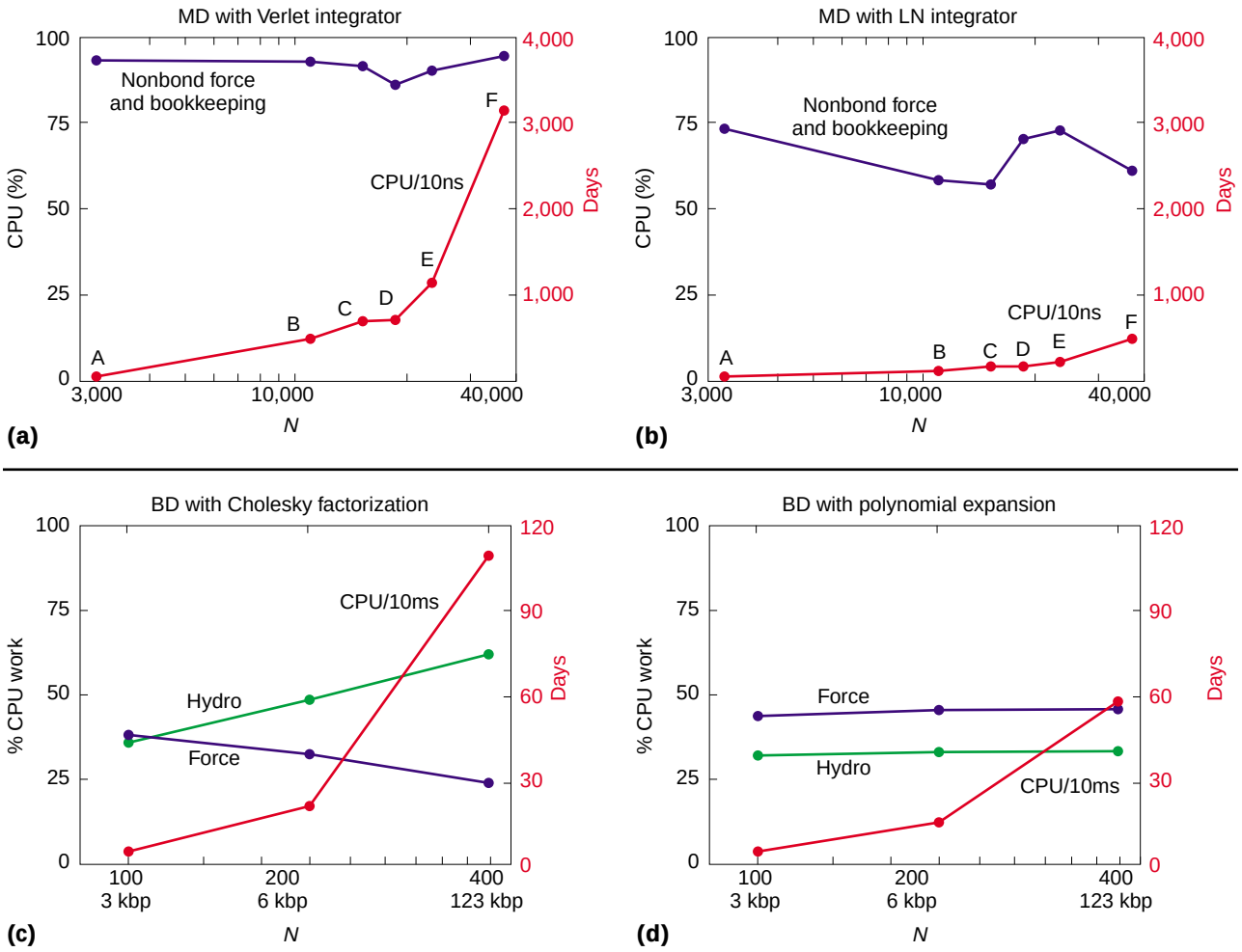

Figure 3. The computational complexity of DNA simulations on the all-atom (top) and macroscopic (bottom) levels. The upper plots correspond to molecular-dynamics calculations, with the fraction of CPU time devoted to calculating the nonbonded energy and forces (blue line) plotted against the number of atoms, *N*. The plot in (a) corresponds to the standard Verlet integrator with Δt = 1.0 fs, while the plot in (b) corresponds to the LN integrator with the time step protocol of 1/2/120 fs. The various all-atom systems correspond to lysozyme (A, 2,857 atoms, *in vacuo*), 12-bp A29 TATA box element (B, 11,013 atoms, solvated in a hexagonal prism), 14-bp A29 TATA box element (C, 15,320 atoms, solvated in a hexagonal prism), triose phosphate isomerase (TIM) (D, 18,733 atoms, solvated in a truncated octahedral prism), TIM (E, 23,635 atoms, solvated in a rectangular prism), and the wildtype TBP/WT DNA complex (F, 37,703 atoms, solvated in a rectangular prism). For the plots in (c) and (d), the fraction of CPU time associated with the hydrodynamics calculations (green lines) and with force calculations (blue line) in Brownian-dynamics simulations of large DNA is plotted versus the system size in kbp. The red curves in all plots correspond to the right-hand axis, the total number of days required to compute a trajectory of 10 ns for all-atom MD, and 10 ms for macroscopic BD. All computations were performed on one R12000 processor of an SGI Origin 2000 with 300-MHz processors.

The above factorization of **D** requires $\mathcal{O}(N^3)$ floating-point operations and consumes most of the CPU's time for BD simulations of large systems. Increasing the efficiency of these hydrodynamics calculations is the key to alleviating the current limitations on the size of DNA systems and the time scale of trajectories that we can simulate using BD.

The Fixman alternative to the Cholesky factorization of **D** involves calculating y, the vector of correlated random numbers, as

$$y = \mathbf{S}z \tag{9}$$

instead of Equation 7. Here **S** is the square root matrix of **D** (**D** = $\mathbf{S}^2$) and not the Cholesky factor **L**. Fixman's idea was to expand the vector y as a

A computational model for supercoiled DNA

Following Stuart Allison's pioneering work,[1] we can represent wormlike DNA as a series of N virtual objects (or beads) connected in a closed loop. The centers of the beads, denoted by r_i, represent a discrete polymer chain's vertices, and local coordinate unit vectors $\{a_i, b_i, c_i\}$ associated with each bead describe the DNA molecule's internal configuration. For circular DNA, the index $i = N + 1$ incides with the first index $i = 1$. Euler angles $\{\alpha_i, \beta_i, \gamma_i\}$ specify the rotation of the $(i - 1)$th to the ith coordinate system.[2]

The configuration-dependent potential energy is modeled as the sum of stretching, bending, twisting, and electrostatic interactions:

$$E = E_S + E_B + E_T + E_C. \quad \text{(A)}$$

We compute the stretching energy E_S from the sum of squared deviations in segment length:

$$E_S = \frac{h}{2}\sum_{i=1}^{N}(\|r_i - r_{i+1}\| - l_o)^2 \quad \text{(B)}$$

where l_o is the resting length of each interbead segment and $l_o = \mathcal{L}_o/N$, where $\mathcal{L}_o$ is the DNA molecule's target length. Setting the stretching constant to $h = 1{,}500 k_B T / l_o^2$ results in deviations in realized segment lengths of less than 1% from l_o.[3,4] We calculate Equation 4 in the main text's discrete analog from the set of Euler angles:

$$E_B + E_T = \frac{A}{2l_o}\sum_{i=1}^{N}\beta_i^2 + \frac{C}{2l_o}\sum_{i=1}^{N}(\alpha_i + \gamma_i - \phi_o)^2. \quad \text{(C)}$$

where ϕ_o is the equilibrium excess twist due to superhelical winding: $\phi_o = 2\pi\sigma(l_o/l_h)$. Here σ is the superhelical density of DNA, $\Delta Lk/Lk_o$ is a normalized linking number difference (typically around –0.05), and l_h is the DNA helical repeat length of about 3.55 nm.

Following Dirk Stigter's work,[5] we approximate the electrostatic energy by the Debye-Hückel potential associated with point charges at the centers of the beads:

$$E_C = \frac{(\nu l_o)^2}{\varepsilon}\sum_{j>i+1}\frac{\exp(-\kappa r_{ij})}{r_{ij}} \quad \text{(D)}$$

where ν is the effective linear-charge density along the chain, ε is the dielectric constant of water, $1/\kappa$ is the Debye length, and r_{ij} is the scalar distance between beads i and j. (We do not consider the $j = i + 1$ term here, because it is counted in the stretching term.) For a monovalent salt concentration of 40 mM, $1/\kappa = 1.52$ nm and $\nu = -3.92$ e•nm^{-1}. (This screening parameter κ should not be confused with the curvature symbol introduced earlier.)

References

1. S.A. Allison, "Brownian Dynamics Simulation of Wormlike Chains: Fluorescence Depolarization and Depolarized Light Scattering," *Macromolecules*, Vol. 19, No. 1, Jan. 1986, pp. 118–124.
2. S.A. Allison, R. Austin, and M. Hogan, "Bending and Twisting Dynamics of Short Linear DNAs: Analysis of the Triplet Anisotropy Decay of a 209 Base Pair Fragment by Brownian Simulation," *J. Chemical Physics*, Vol. 90, No. 7, Apr. 1989, pp. 3843–3854.
3. D. Beard and T. Schlick, "Inertial Stochastic Dynamics: I. Long-Time Step Methods for Langevin Dynamics," *J. Chemical Physics*, Vol. 112, No. 17, May 2000, pp. 7313–7322.
4. D. Beard and T. Schlick, "Inertial Stochastic Dynamics: II. Influence of Inertia on Slow Kinetic Properties of Supercoiled DNA," *J. Chemical Physics*, Vol. 112, No. 17, May 2000, pp. 7323–7338.
5. D. Stigter, "Interactions of Highly Charged Colloidal Cylinders with Applications to Double-Stranded DNA," *Biopolymers*, Vol. 16, No. 7, July 1977, pp. 1435–1448.

series of Chebyshev polynomials, a calculation that requires $\mathcal{O}(N^2)$ operations, compared to the standard method's $\mathcal{O}(N^3)$ operations.

The sidebar "Polynomial expansion for Brownian random force" describes computing the expansion for y. The procedure requires determining bounds on the maximum and minimum eigenvalues of $\mathbf{D}(X^n)$. (In practice, we can approximate these bounds—which are required to scale the matrix for the Chebyshev expansion—by computing the eigenvalues of $\mathbf{D}(X^0)$ and assuming that the magnitudes do not change drastically.) Then, once we have determined the order M of the expansion according to some error criterion, we expand y in terms of polynomials with coefficients determined for the square-root function.

The computational work required for the standard Cholesky treatment of hydrodynamics dominates BD simulations for large systems, as in Figure 3c. When we apply the vector polynomial expansion described earlier, the hydrodynamics calculations consume approximately 33% of the CPU time, regardless of system size. For a 12-kbp system, BD using the standard Cholesky factorization requires twice the CPU time as our implementation of the vector polynomial expansion. This acceleration is not large,

Polynomial expansion for Brownian random force

To expand the vector $y = \mathbf{S}z$, we consider Chebyshev polynomials defined over the interval [-1, 1]. The scaling factors k_1 and k_2 are introduced, where

$$\mathbf{G} = k_1\mathbf{D} + k_2\mathbf{I} \quad \text{(A)}$$

so that the eigenvalues of **G** have magnitudes less than 1. We define the order-M Chebyshev expansion of the square-root matrix as

$$\mathbf{S}_M = \sum_{m=0}^{M} a_m \mathbf{C}_m(\mathbf{G}) \quad \text{(B)}$$

where the $\{a_m\}$ are scalar coefficients and the $\{\mathbf{C}_m\}$ are the Chebyshev polynomial functions **G**. The expansion for $y = \mathbf{S}z$ has a similar form:

$$y_M = \sum_{m=0}^{M} a_m \mathbf{C}_m(\mathbf{G})z = \sum_{m=0}^{M} a_m z_m, \quad \text{(C)}$$

where z_m is the vector $\mathbf{C}_m(\mathbf{G})z$. We found $M = 10$ suitable for our applications to achieve errors of less than 0.1% for 3-kbp systems. (We use a double-precision algorithm, with machine epsilon 10^{-15}.)

We define the Chebyshev polynomials for the matrix expansion according to the formula

$$\mathbf{C}_{m+1} = 2\mathbf{G}\mathbf{C}_m - \mathbf{C}_{m-1};\ \mathbf{C}_0 = \mathbf{I};\ \mathbf{C}_1 = \mathbf{G} \quad \text{(D)}$$

From this, we obtain the polynomials defining the vector expansion for $z_m = \mathbf{C}_m(\mathbf{G})z$ as

$$z_{m+1} = 2k_1\mathbf{D}z_m + 2k_2z_m - z_{m-1};\ z_0 = z;\ z_1 = k_1\mathbf{D}z + k_2z. \quad \text{(E)}$$

Although calculation of $\mathbf{S}_M$ according to Equation B requires a series of matrix–matrix multiplications of complexity $\mathcal{O}(N^3)$, the expansion of y_M defined by Equations C and E involves only matrix–vector multiplications, an $\mathcal{O}(N^2)$ process.

The Chebyshev coefficients for the expansion of a function $g(\lambda)$ are given by

$$a_m = \sum_{j=0}^{M} g(\lambda_j)\, c_m(\lambda_j) \,/\, \|c_m\|^2 \quad \text{(F)}$$

where the λ_j are distributed according to

$$\lambda_j = \cos\left(\frac{2j+1}{M+1}\cdot\frac{\pi}{2}\right). \quad \text{(G)}$$

In Equation F, $c_m(\lambda_j)$ represents the mth Chebyshev polynomial for the scalar case:

$$c_{m+1}(\lambda_j) = 2\lambda_j\, c_m(\lambda_j) - c_{m-1}(\lambda_j);\ c_0(\lambda_j) = 1;\ c_1(\lambda_j) = \lambda_j. \quad \text{(H)}$$

The function $g(\lambda_j)$ is the square root function, scaled by the factors introduced in Equation A:

$$g(\lambda_j) = \left(\frac{\lambda_j - k_2}{k_1}\right)^{1/2}. \quad \text{(I)}$$

We determine the scaling factors k_1 and k_2 so that[1]

$$\begin{aligned} k_1\lambda_{\max} + k_2 &= 1 \\ k_1\lambda_{\min} + k_2 &= -1 \end{aligned} \quad \text{(J)}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are reasonable upper and lower bounds on the eigenvalues of **D**.

Reference

1. M. Fixman, "Construction of Langevin Forces in the Simulation of Hydrodynamic Interaction," *Macromolecules*, Vol. 19, No. 4, 1986, pp. 1204–1207.

because the system size in terms of beads is not large (several hundred, see Table 1). However, we can realize greater CPU gains for larger systems. The Chebyshev alternative to the Cholesky factorization also opens the door to other BD protocols (such as our recent inertial BD idea)[38,39] and is crucial to BD studies of finer models, such as those that are base-pair-based rather than bead-based. Now that the BD computational bottleneck is reduced to electrostatics and hydrodynamics ($\mathcal{O}(N^2)$ for both), fast electrostatic methods help accelerate computation further, especially for the chromatin system,[31] where the number of charges is much greater than the number of hydrodynamic variables or beads.

We have witnessed considerable progress over the past two decades in simulating the dynamics of DNA, both on the all-atom and macroscopic levels.[6,7] It was only in the early 1990s that we could simulate stable, fully solvated models of DNA oligonucleotides with traditional MD methods. Both improved force fields and longer-range electrostatics modeling created these advances. Although such improvements continue, this success has opened the door to investigating many of DNA's subtle sequence-dependent properties that are key to regulatory biological processes. The notion of DNA as a passive partner to protein interactions has largely

been discarded in favor of the view of DNA as an important influencing factor on these processes. Ongoing advances in time-step integration (see Table 1), configurational sampling, and efficient implementation of MD programs on parallel architectures will continue to push the capabilities of DNA and DNA–protein modeling toward experimental time frames. (Of course, these methods are general and also applicable to proteins).

The parallel studies focusing on DNA's structure and kinetics on scales much greater than its persistence length require different algorithmic tools to capture DNA's inherent floppiness and strong dependence on the ionic concentration and solvation. Researchers have applied Monte Carlo and Langevin and Brownian dynamics to these problems, but they encounter computational bottlenecks too. To study kinetic processes of supercoiled DNA, which are largely unresolvable by traditional experimental techniques, these algorithms must be accelerated and broadened in scope. For example, we can replace the traditional $\mathcal{O}(N^3)$ treatment of the random force in BD simulation with a more economical $\mathcal{O}(N^2)$ procedure involving Chebyshev polynomials to allow the study of much larger DNA systems or more refined models where each bead represents a specific base pair. This finer resolution is important for modeling sequence-dependent bending and twisting deformations as observed experimentally (hence appropriate elastic constants can be derived). This enhanced resolution will undoubtedly develop significantly in the next decade. A related review of collective-variable modeling for nucleic acids appears elsewhere.[40]

Ultimately, we must bridge the all-atom and polymer-level representations, but this merging is technically challenging. Hybrid approaches such as those that eliminate the explicit representation of the solvent molecules through the use of generalized Born potentials hold great promise.[41,42] At the spectrum's other end, introducing quantum degrees of freedom through hybrid molecular mechanics–quantum mechanics should broaden the scope of problems that we can study.[43] To be sure, in all these exciting studies, computational scientists will continue to play a key role in advancing our understanding of macromolecular structure and function. CiSE

Acknowledgments

The work on DNA supercoiling started with Wilma Olson, and the recent work on site juxtaposition is in collaboration with Alex Vologodskii. We gratefully acknowledge support from the National Science Foundation (ASC-9157582, ASC-9704681, BIR-9318159), the National Institutes of Health (R01 GM55164), and a John Simon Guggenheim fellowship. Tamar Schlick is an investigator at the Howard Hughes Medical Institute. (See group papers at http://monod.biomath.nyu.edu/.)

References

1. R.R. Sinden, *DNA Structure and Function*, Academic Press, San Diego, Calif., 1994.
2. A.D. Bates and A. Maxwell, "DNA Topology," *In Focus*, Oxford Univ. Press, New York, 1993.
3. D. Crothers, T.E. Haran, and J.G. Nadeau, "Intrinsically Bent DNA," *J. Biological Chemistry*, Vol. 265, No. 13, May 1990, pp. 7093–7096.
4. N.B. Leontis and E. Westhof, "Conserved Geometrical Base-Pairing Patterns in RNA," *Quarterly Rev. Biophysics*, Vol. 31, No. 4, Nov. 1998, pp. 399–455.
5. A.V. Vologodskii and N.R. Cozzarelli, "Conformational and Thermodynamic Properties of Supercoiled DNA," *Ann. Rev. Biophysics Biomolecular Structure*, Vol. 23, 1994, pp. 609–643.
6. D.L. Beveridge and K.J. McConnell, "Nucleic Acids: Theory and Computer Simulation, Y2K," *Current Opinion in Structural Biology*, Vol. 10, No. 2, Apr. 2000, pp. 182–196.
7. W.K. Olson and V.B. Zhurkin, "Modeling DNA Deformations," *Current Opinion in Structural Biology*, Vol. 10, No. 3, June 2000, pp. 286–297.
8. M.A. Young and D.L. Beveridge, "Molecular Dynamics Simulations of an Oligonucleotide Duplex with Adenine Tracts Phased by a Full Helix Turn," *J. Molecular Biology*, Vol. 281, No. 4, Aug. 1998, pp. 675–687.
9. D. Sprous, M.A. Young, and D.L. Beveridge, "Molecular Dynamics Studies of Axis Bending in $d(G_5\text{-}(GA_4T_4C)_2\text{-}C_5)$ and $d(G_5\text{-}(GT_4A_4C)_2\text{-}C_5)$: Effects of Sequence Polarity on DNA Curvature," *J. Molecular Biology*, Vol. 285, No. 4, Jan. 1999, pp. 1623–1632.
10. D. Strahs and T. Schlick, "A-Tract Bending: Insights into Experimental Structures by Computational Models," *J. Molecular Biology*, Vol. 301, No. 3, Aug. 2000, pp. 643–663.
11. G.A. Patikoglou et al., "TATA Element Recognition by the TATA Box-Binding Protein Has Been Conserved Throughout Evolution," *Genes & Development*, Vol. 13, No. 24, Dec. 1999, pp. 3217–3230.
12. G. Guzikevich-Guerstein and Z. Shakked, "A Novel Form of the DNA Double Helix Imposed on the TATA-Box by the TATA-Binding Protein," *Nature Structural Biology*, Vol. 3, No. 1, Jan. 1996, pp. 32–37.
13. L. Pardo et al., "Binding Mechanisms of TATA Box-Binding Proteins: DNA Kinking Is Stabilized by Specific Hydrogen Bonds," *Biophysical J.*, Vol. 78, No. 4, Apr. 2000, pp. 1988–1996.
14. O. Norberto de Souza and R.L. Ornstein, "Inherent DNA Curvature and Flexibility Correlate with TATA Box Functionality," *Biopolymers*, Vol. 46, No. 6, Nov. 1998, pp. 403–441.
15. A. Lebrun, Z. Shakked, and R. Lavery, "Local DNA Stretching Mimics the Distortion Caused by the TATA Box-Binding Protein," *Proc. Nat'l Academy of Science USA*, Vol. 94, No. 4, Apr., 1997, pp. 2993–2998.
16. X. Qian, D. Strahs, and T. Schlick, "Sequence-Dependent Structure and Flexibility of TATA Elements Has Been Selected by the TATA-Box Binding Protein (TBP)," to appear in 2000; schlick@nyu.edu.
17. G. Chirico and J. Langowski, "Brownian Dynamics Simulations of Supercoiled DNA with Bent Sequences," *Biophysical J.*, Vol. 71, No. 2, Aug. 1996, pp. 955–971.
18. D. Sprous and S.C. Harvey, "Action at a Distance in Supercoiled

DNA: Effects of Sequences on Slither, Branching and Intermolecular Concentration," *Biophysical J.*, Vol. 70, No. 4, Apr. 1996, pp. 1893–1908.

19. J.A. Martino, V. Katritch, and W.K. Olson, "Influence of Nucleosome Structure on the Three-Dimensional Folding of Idealized Minichromosomes," *Structure with Folding & Design*, Vol. 7, No. 8, Aug. 1999, pp. 1009–1022.
20. L. Ehrlich et al., "A Brownian Dynamics Model for the Chromatin Fiber," *Computer Applications in the Biosciences*, Vol. 13, No. 3, June 1997, pp. 271–279.
21. K. Luger et al., "Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution," *Nature*, Vol. 389, No. 6648, Sept. 1997, pp. 251–260.
22. T. Schlick, "Modeling Superhelical DNA: Recent Analytical and Dynamic Approaches," *Current Opinion in Structural Biology*, Vol. 5, No. 2, Apr. 1995, pp. 245–262.
23. T. Schlick, E. Barth, and M. Mandziuk, "Biomolecular Dynamics at Long Time Steps: Bridging the Timescale Gap between Simulation and Experimentation," *Ann. Rev. Biophysics Biomolecular Structure*, Vol. 26, 1997, pp. 179–220.
24. T. Schlick et al., "Algorithmic Challenges in Computational Molecular Biophysics," *J. Computational Physics*, Vol. 151, No. 1, May 1999, pp. 9–48.
25. P. Koehl and M. Levitt, "Theory and Simulation: Can Theory Challenge Experiment?" *Current Opinion in Structural Biology*, Vol. 9, No. 2, Apr. 1999, pp. 155–156.
26. S. Doniach and P. Eastman, "Protein Dynamics Simulations from Nanoseconds to Microseconds," *Current Opinion in Structural Biology*, Vol. 9, No. 2, Apr. 1999, pp. 157–163.
27. E. Barth and T. Schlick, "Overcoming Stability Limitations in Biomolecular Dynamics: I. Combining Force Splitting via Extrapolation with Langevin Dynamics in LN," *J. Chemical Physics*, Vol. 109, No. 5, Aug. 1998, pp. 1617–1632.
28. E. Barth and T. Schlick, "Extrapolation versus Impulse in Multiple-Time Stepping Schemes: II. Linear Analysis and Applications to Newtonian and Langevin Dynamics," *J. Chemical Physics*, Vol. 109, No. 5, Aug. 1998, pp. 1632–1642.
29. Y. Duan and P.A. Kollman, "Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution," *Science*, Vol. 282, No. 5389, 23 Oct. 1998, pp. 740–744.
30. D.L. Ermak and J.A. McCammon, "Brownian Dynamics with Hydrodynamic Interactions," *J. Chemical Physics*, Vol. 69, No. 4, Aug. 1978, pp. 1352–1360.
31. D. Beard and T. Schlick, "Computational Modeling Predicts the Structure and Dynamics of the Chromatin Fiber," submitted to *Structure with Folding & Design*, 2000; schlick@nyu.edu.
32. M. Fixman, "Construction of Langevin Forces in the Simulation of Hydrodynamic Interaction," *Macromolecules*, Vol. 19, No. 4 1986, pp. 1204–1207.
33. R.M. Jendrejack, M.D. Graham, and J.J. de Pablo, "Hydrodynamic Interactions in Long Chain Polymers: Application of the Chebyshev Polynomial Approximation in Stochastic Simulations," *J. Chemical Physics*, Vol. 113, No. 7, Aug. 2000, pp. 2894–2900.
34. M. Kröger et al., "Variance Reduced Brownian Simulation of a Bead-Spring Chain under Steady Shear Flow Considering Hydrodynamic Interaction Effects," *J. Chemical Physics*, Vol. 113, No. 11, Sept. 2000, pp. 4767–4773.
35. P.J. Hagerman, "Flexibility of DNA," *Ann. Rev. Biophysics Biophysical Chemistry*, Vol. 17, 1988, pp. 265–286.
36. H. Jian, T. Schlick, and A. Vologodskii, "Internal Motion of Supercoiled DNA: Brownian Dynamics Simulations of Site Juxtaposition," *J. Molecular Biology*, Vol. 284, No. 2, Nov. 1998, pp. 287–296.
37. J. Rotne and S. Prager, "Variational Treatment of Hydrodynamic Interaction in Polymers," *J. Chemical Physics*, Vol. 50, 1969, pp. 4831–4837.
38. D. Beard and T. Schlick, "Inertial Stochastic Dynamics: I. Long-Time Step Methods for Langevin Dynamics," *J. Chemical Physics*, Vol. 112, No. 17, May 2000, pp. 7313–7322.
39. D. Beard and T. Schlick, "Inertial Stochastic Dynamics: II. Influence of Inertia on Slow Kinetic Properties of Supercoiled DNA," *J. Chemical Physics*, Vol. 112, No. 17, May 2000, pp. 7323–7338.
40. I. Lafontaine and R. Lavery, "Collective Variable Modeling of Nucleic Acids," *Current Opinion in Structural Biology*, Vol. 9, No. 2, Apr. 1999, pp. 170–176.
41. B.N. Dominy and C.L. Brooks, III, "Development of a Generalized Born Model Parameterization for Proteins and Nucleic Acids," *J. Physical Chemistry B*, Vol. 103, No. 18, May 1999, pp. 3765–3773.
42. D. Bashford and D.A. Case, "Generalized Born Models of Macromolecular Solvation Effects," *Ann. Rev. Physical Chemistry*, Vol. 51, 2000, pp. 129–152.
43. J. Gao, "Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials," *Reviews in Computational Chemistry*, Vol. 7, K.B. Lipkowitz and D.B. Boyd, eds., VCH Publishers, New York, 1996, pp. 119–185.

Tamar Schlick is a professor of chemistry, mathematics, and computer science at New York University. She is also an associate investigator at the Howard Hughes Medical Institute. Her technical interests include computational and structural biology, specifically on algorithms for biomolecular modeling and simulations and their application to proteins and nucleic acids. She received her PhD in mathematics from the Courant Institute of Mathematical Sciences. Contact her at the Dept. of Chemistry and Courant Inst. of Mathematical Sciences, New York Univ., 251 Mercer St., New York, NY 10012; schlick@nyu.edu.

Daniel A. Beard is a postdoctoral fellow at New York University. He received his PhD in bioengineering from the University of Washington. Contact him at the Dept. of Chemistry, 31 Washington Place, 1021 Main, New York Univ., New York, NY 10003; beard@biomath.nyu.edu.

Jing Huang is a fourth-year graduate student in chemistry at New York University. Contact her at the Dept. of Chemistry, 31 Washington Place, 1021 Main, New York Univ., New York, NY 10003; jingh@biomath.nyu.edu.

Daniel A. Strahs is a Howard Hughes Medical Institute research specialist at New York University. He received his PhD in biochemistry from the Albert Einstein College of Medicine. Contact him at the Dept. of Chemistry, 31 Washington Place, 1021 Main, New York Univ., New York, NY 10003; dan.strahs@nyu.edu.

Xiaoliang Qian is a fifth-year graduate student in chemistry at New York University. Contact him at the Dept. of Chemistry, 31 Washington Place, 1021 Main, New York Univ., New York, NY 10003; qian@biomath.nyu.edu.