

From Atomic to Mesoscopic Descriptions of the Internal Dynamics of DNA

N. Bruant,* D. Flatters,# R. Lavery,# and D. Genest*

*Centre de Biophysique Moléculaire, CNRS UPR 4301 and the University of Orléans, 45071 Orléans, and #Laboratoire de Biochimie Théorique, CNRS UPR 9080, Institut de Biologie Physico-Chimique, 75005 Paris, France

ABSTRACT An analysis of four 1-ns molecular dynamics trajectories for two different 15-bp oligonucleotides is presented. Our aim is to show which groups of atoms can be treated as rigid bodies within a bead representation of DNA, independently of the base sequence and for any conformations belonging to the A/B family. Five models with moderate intragroup deformations are proposed in which the groups are formed of atoms belonging to a single nucleotide or to a complementary nucleotide pair. The influence of group deformation in two of these models is studied using canonical correlation analysis, and it is shown that the internal DNA dynamics is indeed dominated by the rigid motion of the defined atom groups. Finally, using one of the models within a bead representation of duplex DNA makes it possible to obtain stretching, torsional, and bending rigidities in reasonable agreement with experiment but points to strongly correlated stretching motions.

INTRODUCTION

The flexibility and internal dynamics of DNA are believed to play a major role in specific protein-DNA recognition. However, while many experiments measuring rotational correlation times have provided evidence of DNA deformations covering nanosecond (Wahl et al., 1970; Genest and Wahl, 1978; Hogan and Jardetzsky, 1980; Millard et al., 1988) to millisecond (Leroy et al., 1985; Guéron et al., 1987) time scales, they have not provided an overall spatial description of the internal motions.

An alternative approach to such data is offered by molecular modeling techniques such as molecular mechanics (Lavery, 1994; Packer and Hunter, 1998), Monte Carlo (Zhurkin et al., 1991; Gabb et al., 1997) or molecular dynamics (MD) (Sproun et al., 1998; Feig and Pettitt, 1998) simulations. Dynamic methods are able to describe structural fluctuations and transitions at the atomic level, as a function of base sequence, while taking solvent and counterion effects explicitly into account. However, such calculations are very time consuming, and only short oligonucleotides (~20–30 bp) can currently be simulated for periods of a few nanoseconds. For many purposes, such data are insufficient, and it is necessary to envisage models of larger DNA fragments and simulations covering much longer periods. On the other hand, it is clearly possible to obtain useful information on nucleic acids without always requiring atomically resolved detail. Both of these facts suggest the importance of developing lower resolution models.

At a mesoscopic level, different simplified physical models have already been developed, in which double-stranded DNA is treated as a long flexible filament, with the environment reduced to simple continuum or stochastic effects (Schlick, 1995; Olson, 1996; Lafontaine and Lavery, 1999). These models may be classed in two broad categories: continuous elastic rods (Barkley and Zimm, 1979; Schlick and Olson, 1992) or strings of beads (Allison and McCammon, 1984; Chirico and Langowski, 1994; Jian et al., 1998; Tan and Harvey, 1989). They implicitly correspond to freezing many degrees of freedom, but which degrees of freedom are frozen is a choice generally based on a priori hypotheses, which are not fully under control. As an example, recent work by Chirico and Langowski (1994) uses a bead model in which each bead is chosen to correspond to a rigid 37-bp DNA segment. Such models are generally parameterized on the basis of a small number of experimental data, such as persistence length measurements or fluorescence anisotropy decay. However, these data are often imprecise and do not cover finer effects, most notably those related to base sequence. They are also only valid for relatively small deformations, which involve neither local nor global structural transitions. Despite these restrictions, bead and rod models of DNA are very important, at least because they allow long simulations on large systems with a very limited number of variables. Dynamic simulations on fragments with thousands of base pairs over millisecond time scales thus become possible.

The present paper is an attempt to establish a bridge between atomic and mesoscopic descriptions of the internal dynamics of DNA and, in particular, to make judicious choices for freezing degrees of freedom. The principle is to use MD trajectories to determine which groups of atoms move collectively and can thus be frozen and employed in a well-founded rigid-body model of DNA.

We have recently developed methods for detecting rigid-body motions within MD trajectories (Gaudin et al., 1997; Hery et al., 1998), which also allow us to study what effects

Received for publication 21 December 1998 and in final form 2 July 1999.

Address reprint requests to Dr. Daniel Genest, Centre de Biophysique Moléculaire, CNRS, rue Charles Sadron, 45071 Orleans cedex 2, France. Tel.: 33-238-25-55-93; Fax: 33-238-63-15-17; E-mail: genest@cns-orleans.fr.

Dr. Flatters's present address is Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England.

© 1999 by the Biophysical Society

0006-3495/99/11/2366/11 \$2.00

neglecting the residual deformations of these bodies have on the internal dynamics of a molecule (Genest, 1996, 1998; Briki and Genest, 1995). These methods have been tested on two different double-stranded DNA sequences, an octanucleotide and a dodecanucleotide. It was found that each nucleotide could be satisfactorily modeled by three rigid bodies: the base, the sugar ring, and an extended phosphate group ($\text{PO}_4 + \text{C5}'$). However, the generality of these results was limited because 1) the MD simulations were short (200–250 ps), 2) they were restricted to fluctuations around the B-DNA conformation, and 3) they were limited to a single force field (GROMOS 87; van Gunsteren and Berendsen, 1986). It was also noted that although only one set of rigid bodies was tested, other solutions were possible.

In this study, we present a much more thorough investigation for two new 15-bp sequences. The first one, GCGTATATAAAACGC, includes a strong binding site (TATATAA) for the TATA box binding protein (TBP), and the second one, GCGTAAAAAAACGC (with two T \rightarrow A mutations underlined), includes a weak binding site. Each sequence was simulated twice, using different initial structures, and each simulation lasted 1 ns (Flatters et al., 1998; Flatters and Lavery, 1998). The simulations were performed with AMBER, using the Parm94 force field (Pearlman et al., 1995; Cornell et al., 1995), and with particle mesh Ewald summations to avoid short-range electrostatic cutoffs (Darden et al., 1993; Cheatham et al., 1995). During these simulations A \leftrightarrow B transitions occurred for the central segment of the oligonucleotides.

This much more extensive data set allows us to overcome the restrictions applying to our earlier analysis and to define a hierarchy of possible rigid body models with known accuracy. We use one of the resulting models to construct a bead representation of the oligomers studied and to extract the torsional, bending, and stretching rigidities, which can be compared with experiment. We also use the sequence differences between the two oligomers to look at their impact on the rigid-body models.

METHODS

Molecular dynamics simulations

Four 1-ns trajectories are analyzed, corresponding to two different 15-bp DNA oligonucleotides (1: GCGTATATAAAACGC, 2: GCGTAAAAAAACGC), each simulated twice with different initial configurations: 1) all sugars have B-like C2'-endo pucker and 2) with A-like C3'-endo pucker for eight central base pairs (with the exception of the 3'-termini) in line with the conformations observed in the TBP-DNA complex (Kim et al., 1993). These two different initial conditions are referred to as B and BAB, respectively. An analysis of these four simulations, relating dynamic behavior to their known propensity for binding TBP, has recently been reported (Flatters et al., 1997; Flatters and Lavery, 1998).

All simulations were performed with the AMBER 4.1 program (Pearlman et al., 1995), using the Parm94 all-atom force field (Cornell et al., 1995). Each simulation was carried out in a box of ~ 4500 water molecules, and electrostatic neutrality was achieved by adding 28 Na^+ ions. Electrostatic interactions were treated with a particle mesh Ewald summation (Darden et al., 1993; Cheatham et al., 1995). A 2-fs time step was used in conjunction with SHAKE constraints (Ryckaert et al., 1977) on all bond lengths involving hydrogen atoms, and overall translations and rotations of DNA were removed. Configurations were stored every 0.5 ps, leading to a total of 1900 configurations for each trajectory.

Search for groups of atoms

The principle of searching for quasirigid atom groups is based on an analysis of the matrix of the root mean square fluctuations (RMSs) of interatomic distances (limited in this analysis to all nonhydrogen atoms). Thus, two atoms, i and j , belong to the same rigid body if the RMS of separation r_{ij} ($\langle r_{ij}^2 \rangle^{1/2}$) is smaller than a fixed tolerance r_c . Choosing r_c makes it possible to build a Boolean decision matrix D for which each element D_{ij} is 1 if $\langle r_{ij}^2 \rangle^{1/2} < r_c$ and 0 otherwise. Different procedures can then be used to determine which atoms should be grouped (Gaudin et al., 1997). This leads to a reordered D matrix, which can be displayed graphically.

To maintain the notion of base sequence, we currently focus on groups for which all of the atoms belong to a single nucleotide or to a complementary nucleotide pair. For generality, we accept only those groups involving equivalent atoms for each nucleotide (or complementary nucleotide pair) within all of the trajectories studied.

Atom group dynamics

Let $\{F_m\}$ be a frame bound to the molecule that is invariant over the time course of a simulation, inasmuch as tumbling and translational motions are suppressed. Let us consider the first configuration of the trajectory, and let $\{F_g^1\}$ be the frame defined by the principal axes of a chosen group of atoms. The position of the center of mass and the orientation of the axes are defined relative to $\{F_m\}$. The first configuration of this group is now successively fitted to the other configurations k of this group, using the superposition method of McLachlan (1979). This corresponds to a translation of the center of mass of the rigid body and a rotation defined by three Euler angles, moving $\{F_g^1\}$ into $\{F_g^k\}$. The coordinates of each atom in the group at configuration k can then be computed with respect to $\{F_g^k\}$, thereby defining the deformation of the group. Time series of these coordinates are also straightforward to calculate. We define the time series of the center of mass of a group within the molecular frame as

$$\text{DEV}_{\text{CM}}(k) = [x_k^2 + y_k^2 + z_k^2]^{1/2} \quad (1a)$$

and of the mean atomic position within the rigid body frame as

$$\text{DEV}_{\text{at}}(k) = [(1/N_a) \sum_a (x_{k,a}^2 + y_{k,a}^2 + z_{k,a}^2)]^{1/2} \quad (1b)$$

In these expressions, N_a is the number of atoms of a group. The indices k and a correspond, respectively, to the k th configuration and the atom label in the group. The sum in Eq. 1b is over the atoms of the group ($a = 1, \dots, N_a$). x , y , and z are the coordinates of the center of mass defined in $\{F_m\}$ (Eq. 1a) or of the atoms of the group defined in $\{F_g^k\}$ (Eq. 1b).

The corresponding root mean square fluctuations are given by

$$\text{RMS}_{\text{CM}} = \left\{ \frac{1}{N} \sum_k [(x_k - \langle x \rangle)^2 + (y_k - \langle y \rangle)^2 + (z_k - \langle z \rangle)^2] \right\}^{1/2} \quad (2a)$$

and

$$\text{RMS}_{\text{at}} = \left[\frac{1}{N} \frac{1}{N_a} \sum_k \sum_a \{ (x_{k,a} - \langle x_a \rangle)^2 + (y_{k,a} - \langle y_a \rangle)^2 + (z_{k,a} - \langle z_a \rangle)^2 \} \right]^{1/2} \quad (2b)$$

Equations 1a and 2a are related to the translation of the groups, while Eqs. 1b and 2b are related to their deformation.

Canonical correlation analysis

As explained in the previous paragraph, the motion of each group of atoms may be decomposed into a translation, a rotation, and a deformation. The approximation of rigid groups is fully valid only if the deformation of any group and the rigid-body motions of other groups are uncorrelated. A method for quantifying this has recently been described (Briki and Genest, 1994, 1995; Genest, 1996). Briefly, let n and m (we assume $n \leq m$) be the number of coordinates describing the various motions of two groups composed of N_a and N'_a atoms, respectively (note, $n = m = 3$ for translation and rotation, while $n = 3N_a$ and $m = 3N'_a$ for deformations). These two sets of coordinates can be considered as N -dimensional vectors (with N equal to the number of configurations) that define, respectively, n - and m -dimensional subspaces. The correlation between the two groups of vectors is related to the relative orientations of the two subspaces. Let R11 be the correlation matrix for the first set of vectors, let R22 be the correlation matrix for the second set, let R12 be the correlation matrix between vectors of the first group and the second group, and let R21 be the transpose of R12. We further define a square symmetrical positive definite matrix $[R] = [R11]^{-1} \cdot [R12] \cdot [R22]^{-1} \cdot [R21]$. A canonical correlation coefficient can be defined that is related to the trace (Tr) of $[R]$ by $M =$

$\{(1/n)\text{Tr}(\mathbf{R})\}^{1/2}$ (Briki and Genest, 1994). In practice, the n and m vectors are not the actual coordinates, but their normalized deviation from the corresponding mean. If the components of the n vectors are taken at the same time as those of the m vectors for calculating the correlation matrices, one gets an equal time correlation coefficient, whereas if a time delay is introduced between the two sets of components, a time correlation function may be calculated as a function of the delay. Similarly, comparing one set of vectors with itself allows autocorrelation functions to be computed, while using two different sets of vectors leads to cross-correlation functions.

Parameters for the bead model

Let us consider two consecutive base pairs to be represented by two rigid beads linked by a virtual bond between their centers of mass C1 and C2. Let $\{\mathbf{u1}, \mathbf{v1}, \mathbf{w1}\}$ and $\{\mathbf{u2}, \mathbf{v2}, \mathbf{w2}\}$ be the principal axes of each bead calculated at the first step of the simulation. Owing to the shape of the base pairs, the two first axes of each bead lie roughly in the mean planes of the base pairs, while the third is perpendicular to these planes. As a consequence of rigid body motions, C1 and C2 translate and the axes rotate as a function of time. We define a local bending by the angle β between $\mathbf{w1}$ and $\mathbf{w2}$, a local twisting by the angle τ between the planes $(\mathbf{C1C2}, \mathbf{u1})$ and $(\mathbf{C1C2}, \mathbf{u2})$, and a local stretching by the distance $l = |\mathbf{C1C2}|$. According to an earlier bead model (Chirico and Langowski, 1994; Klenin et al., 1998) the associated elastic potentials are

$$U_\beta = (1/2)(B/l_0)(\beta - \beta_0)^2 \text{ for bending} \quad (3a)$$

$$U_\tau = (1/2)(C/l_0)(\tau - \tau_0)^2 \text{ for twisting} \quad (3b)$$

$$U_l = (1/2)(S/l_0)(l - l_0)^2 \text{ for stretching} \quad (3c)$$

where B , C , and S are the local bending, twisting, and stretching rigidities at the levels of the different beads. β_0 , τ_0 , l_0 are the equilibrium values of β , τ , and l , respectively. Relationships exist between B , C , and S and the RMS fluctuations $\langle(\Delta\beta)^2\rangle^{1/2}$, $\langle(\Delta\tau)^2\rangle^{1/2}$ and $\langle(\Delta l)^2\rangle^{1/2}$ of β , τ , and l , respectively, on the other hand:

$$B = 2l_0kT/\langle(\Delta\beta)^2\rangle \quad (4a)$$

$$C = l_0kT/\langle(\Delta\tau)^2\rangle \quad (4b)$$

$$S = l_0kT/\langle(\Delta l)^2\rangle \quad (4c)$$

The equilibrium values and the RMS fluctuations can easily be determined from the time series of the center of mass and the principal axes of the beads. The B , C , and S values that are presented in this work correspond to averages along the oligomers studied.

For S a global value has also been calculated according to two different procedures. In the first, l_0 and $\langle(\Delta l)^2\rangle$ in Eq. 4c are related to the distance between the first and last beads of the sequence, while, in the second, they are related to the

sum of the distances between consecutive beads, which is in fact the length of the DNA sequence.

RESULTS

Models for rigid-body description

Following the value chosen for the tolerance factor r_c , different rigid-body models of DNA are generated. The largest root mean square fluctuation among the full set of interatomic distances is $\sim 0.45 \pm 0.05$ nm for each of the four trajectories. Setting r_c to this value thus assimilates the entire oligomers as single rigid bodies. Decreasing r_c to 0.15 nm allows us to extract six rigid bodies, each composed of several nucleotides, either on the same strand or on different strands of the oligomers. However, these intermediate level rigid bodies are not well defined, as different results are obtained from the different trajectories studied. Decreasing r_c to 0.1 nm increases the number of rigid bodies but does not improve their definition.

We therefore turned to a systematic analysis of each nucleotide and of each complementary nucleotide pair, to look for general rules governing the definition of rigid groups of atoms. In this way, five different rigid-body models (see Fig. 1 and Table 1) can be constructed as a function of r_c . For a very small r_c (≤ 0.012 nm), every nucleotide can be described by four rigid groups: the base, the sugar ring, the phosphate, and C5' (model 1). When r_c reaches 0.020 nm, the C5' atoms can be regrouped with the phosphate (model 2). For $r_c = 0.030$ nm, paired bases can be grouped together, while the sugars and the $\text{PO}_4 + \text{C5}'$ groups remain separated, leading to a total of five rigid bodies for each complementary nucleotide pair (model 4). For $r_c = 0.044$ nm, it is found that each sugar can be grouped with its associated base, so that each nucleotide is defined by only two rigid bodies (model 3). Finally, for $r_c = 0.07$ nm the sugar-base groups of a nucleotide pair can be coalesced, leaving only two backbone $\text{PO}_4 + \text{C5}'$ groups

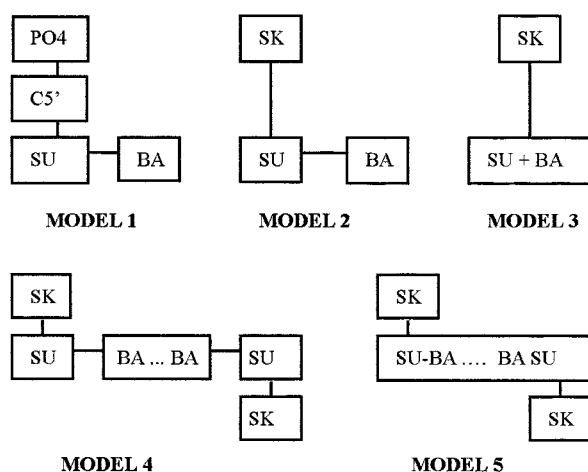


FIGURE 1 Five models for quasirigid groups of atoms according to tolerance r_c . SK, $\text{PO}_4 + \text{C5}'$; SU, sugar ring; BA, base.

TABLE 1 Maximum interatomic distance (r_c) RMS fluctuations (nm) of the groups for the different models and simulations

Trajectory	Model 1	Model 2	Model 3	Model 4	Model 5
1-B	0.010	0.018	0.044	0.030	0.065
1-BAB	0.010	0.015	0.042	0.028	0.060
2-B	0.010	0.015	0.026	0.028	0.070
2-BAB	0.012	0.020	0.044	0.028	0.070

(model 5). At this level, it is still impossible to group together all atoms belonging to two consecutive nucleotides. In addition, it is only possible to group together all of the atoms of any given nucleotide when $r_c > 0.12$ nm.

Atom group dynamics

For each MD trajectory, the time series of the center-of-mass coordinates and of the Euler angles defining the position and the orientation of each group frame $\{F_g^k\}$ (see Methods) were calculated with respect to the oligonucleotide-bound axis system $\{F_m\}$. These time series describe the rigid-body motions of each group within the molecule. The time series of the atomic coordinates relative to $\{F_g^k\}$ of each group were also computed, describing the internal deformation of the group during the simulation. It is therefore possible to compare the relative amplitudes of the rigid body and internal fluctuations of any given group for a particular trajectory. The amplitudes are defined by the RMS fluctuations of the center of mass and the mean atomic RMS fluctuation of the deformation (Eqs. 2a and 2b). These RMS values, averaged over all equivalent groups, are presented in Tables 2 and 3 for each trajectory. The groups corresponding to single bases are excluded because of their very small deformations. It can be seen from these results that the groups composed of a single sugar ring or of an extended phosphate group ($\text{PO}_4 + \text{C5}'$) in models 2 and 4 undergo translational fluctuations that are at least an order of magnitude higher than their deformation fluctuations. For the other groups this ratio is reduced by at least a factor of 2.

In the case of models 2 and 4, which have particularly small deformations, we have computed the time series of the mean atomic position in the frame bound to the groups (Eq. 1b) and of their center of mass within the molecular frame (Eq. 1a). Typical examples are given in Fig. 2. Two types of behavior are observed. Some time series show a few distinct conformational transitions, while others exhibit fluctuations around an average conformation or position. This second type of time series was Fourier transformed, leading to the data shown in Fig. 3. It is seen that low frequencies are only strongly represented in the position spectrum.

Correlated motions

To avoid end effects, only the nine central base pairs of the oligomers have been analyzed, and comparisons are limited to models 2 and 4 (the minor deformations of the bases in

TABLE 2 Center-of-mass fluctuations RMS_{CM} in nm (Eq. 2a) of the different groups within the molecular reference frame

Models	1, 2, 4	2, 3, 4, 5	3	4	5
Groups	SU	SK	SU + BA	BA - BA	SU - BA ... BA - SU
1-B	0.090	0.110	0.076	0.068	0.068
1-BAB	0.073	0.096	0.061	0.055	0.070
2-B	0.096	0.116	0.084	0.073	0.075
2-BAB	0.078	0.098	0.066	0.060	0.060

model 2 are ignored). As mentioned above, the time series of certain groups exhibits a small number of transitions that do not allow statistically accurate correlation coefficients to be obtained. In such cases it is still possible to detect correlations qualitatively by a visual comparison of the corresponding time series. An example is given in Fig. 4 for the translational motion of two adjacent sugar rings. However, such cases concern less than 30% of all nucleotides.

We first note that the equal time translational correlation between different groups is large. The correlation decreases only slowly as the distance between the groups increases. An average value of 0.88 is obtained for covalently linked groups, against 0.50 for groups separated by six nucleotides. Rotational motions show a similar behavior, although the values are significantly lower (0.50 for linked groups and 0.25 for distant groups).

In sharp contrast, there is very little correlation between the deformations of different groups, whatever their separation (~ 0.15). These findings are in good agreement with previous studies (Briki and Genest, 1995; Genest, 1996). No significant differences are found between nucleotides, between the two oligonucleotide sequence, or between the trajectories.

It is also interesting to note that the deformations of the sugars or of the extended phosphate groups are only very weakly correlated with the translational and rotational motions of other groups (< 0.25) and consequently have no significant effect on the rigid-body motions of other groups. This result holds even between a sugar and its associated base (model 2) or base pair (model 4).

A somewhat different result is found for the equal time correlation between the deformation of a sugar or of an extended phosphate group and its own rigid-body motions. Although no effect is seen on translation, a correlation of $\sim 0.4-0.5$ is found between deformation and rotation. Equal time correlation coefficients between the deformation of a base pair in model 4 and rigid-body motions of other base

pairs also reveal a small correlation (0.25–0.33), but a problem of accuracy may occur in this case (see below).

Examples of canonical cross-correlation functions between group deformations and the deformation or rigid-body motions of other groups are shown over 100 ps in Fig. 5. It can be seen that a given group deformation has no effect on either the deformations or the rigid-body motions of other groups, even after a delay of 100 ps. This is true for any sugar or extended phosphate groups of models 2 and 4. However, the rigid-body motions of different groups are found to be correlated over a time period of ~ 40 ps (Fig. 6).

Relationship with mesoscopic models

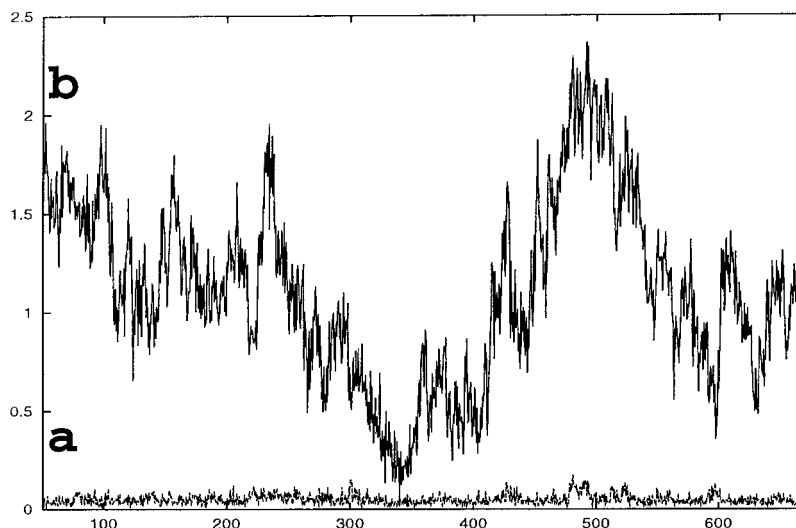
This analysis has been performed with model 4 to formally reproduce the bead models used by other authors (Allison and McCammon, 1984; Chirico and Langowski, 1994; Kleinin et al., 1998). It is again limited to the nine central base pairs to avoid oligomeric end effects. Each base pair is assumed to be a single rigid bead. The sugar and extended phosphate groups play the role of springs linking the beads. The configuration averaged values and the corresponding RMS fluctuations of local twist and bending angles and of the distances between consecutive beads were calculated for each base pair and each trajectory. Torsional, bending, and stretching rigidities were then evaluated. Although some differences are observed between beads, it is not possible to assign these to sequence or trajectory-linked effects in any reliable way. The bead-averaged values are given in Table 4. These values are on the same order of magnitude as those obtained experimentally, with the exception of the stretching rigidity (mean value = 3760 pN), which is significantly higher than values given by Cluzel et al. (1996) and by Smith et al. (1996), which are on the order of 900–1100 pN.

Because experimental measurements of DNA stretching give global information on long, single molecules and not local information at the level of individual base pairs, we

TABLE 3 Average atomic fluctuations RMS_{at} in nm (Eq. 2b) of the different groups within a group-bound reference frame

Models	1, 2, 4	2, 3, 4, 5	3	4	5
Groups	SU	SK	SU + BA	BA - BA	SU - BA ... BA - SU
1-B	0.0062	0.0048	0.011	0.013	0.020
1-BAB	0.0067	0.0047	0.011	0.014	0.021
2-B	0.0055	0.0048	0.010	0.012	0.017
2-BAB	0.0073	0.0049	0.013	0.013	0.019

FIGURE 2 Time evolution of the sugar ring T25 in simulation 1-B. (a) $DEV_{at}(k)$: mean atomic position within the group frame (deformation). (b) $DEV_{CM}(k)$: center of mass position within the molecular frame (translation).

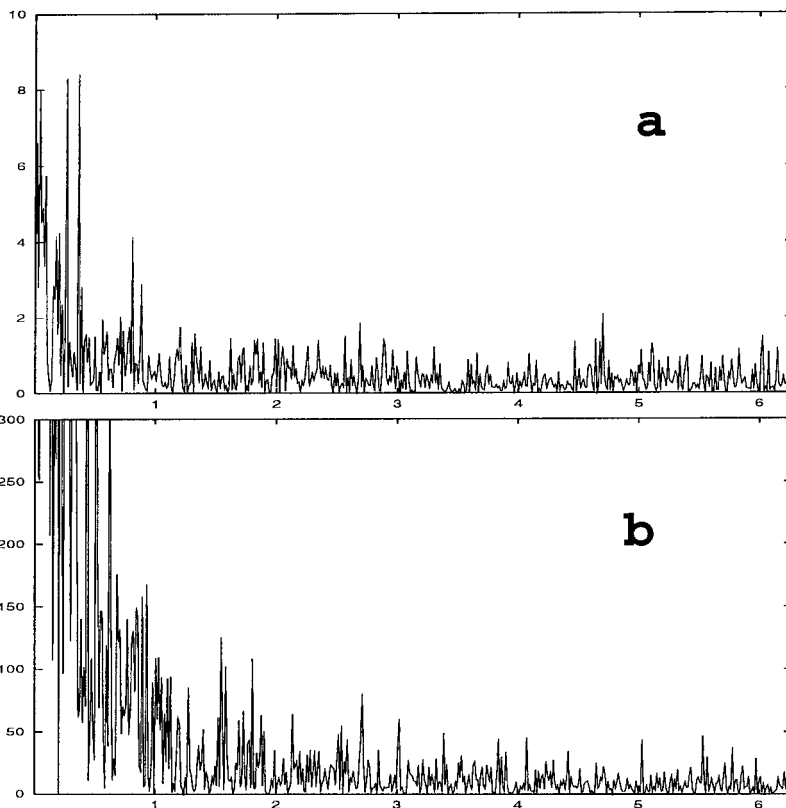


have used two other procedures for calculating a more macroscopic S value (see Methods). In the first, the average distance and the corresponding RMS fluctuations are measured between the two separated base pairs A4 and T12, while in the second, the sum of the interbead distances lying between A4 and T12 is calculated for each conformation of the simulation, and its average and RMS fluctuations are used to determine S . We find that these two procedures lead to a decrease in S , by a factor of 2.5 in the first case (mean value 1530 pN) and by a factor of 1.7 in the second case (mean value 2260 pN).

DISCUSSION AND CONCLUSION

This study provides a quantitative justification for five different rigid-body models of DNA. The models that represent each nucleotide or complementary nucleotide pair by a few beads are valid for oligomers with different sequences and for both A- and B-like helical conformations. These beads are closely related to chemical entities within the DNA molecule and can be expected to be valid for other base sequences. Each model has a defined degree of accuracy based on interatomic distance fluctuations. Models 1

FIGURE 3 Fourier transforms of the time series shown in Fig. 2 for deformation (a) and translation (b).



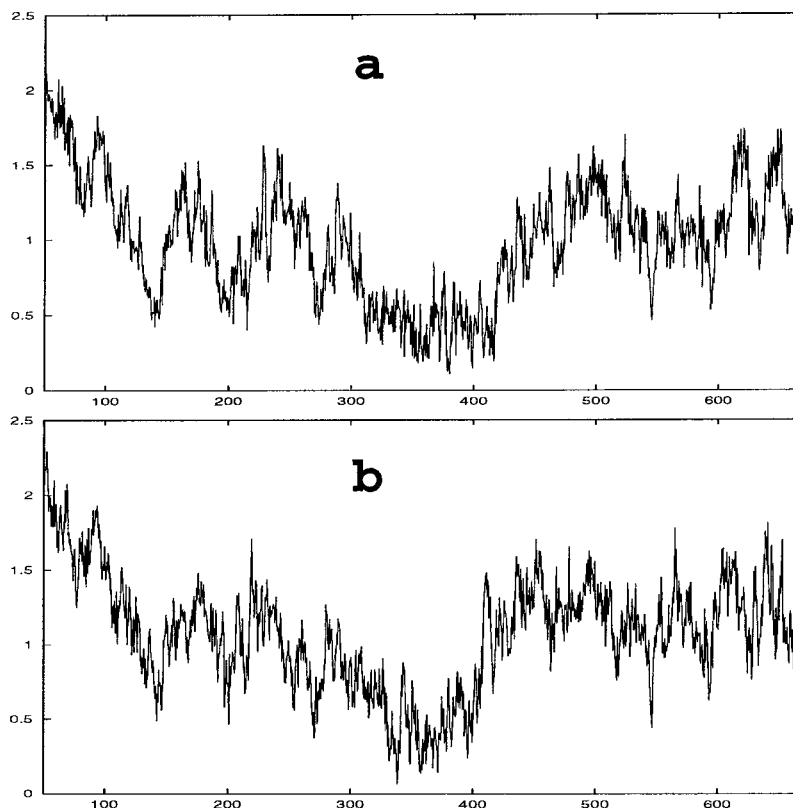


FIGURE 4 Time series of the center of mass positions ($DEV_{CM}(k)$) of the sugar rings of nucleotides A5 (a) and A6 (b) in simulation 1-B.

and 2 defined here are identical to those reported earlier on the basis of two short MD simulations (200–250 ps) limited to B-DNA (Gaudin et al., 1997) and using a different force field. The present work increases the generality of these definitions and lays a firm foundation for the formulation of mesoscopic models of DNA, lying in the relatively unexplored area between atomic and elastic rod representations, and enabling sequence-dependent effects to be maintained.

The rigid-body models we define can be divided into two categories: 1) models 1, 2 and 3, where beads contain atoms belonging to a single nucleotide, and 2) models 4 and 5, where certain beads contain atoms from complementary nucleotides. The precision of the models decreases in the order 1, 2, 4, 3, 5. In all models, the bases, the sugar, and the phosphate group can be treated as rigid. Although it is possible in some models to assimilate a sugar and a base or a base pair within a single rigid body, the phosphate group bead cannot be extended beyond the adjacent C5' atom.

For models 2 and 4 we have used canonical correlation analysis to determine whether freezing internal degrees of freedom within a bead can influence the rigid-body motions. These two models share backbones divided into sugars and extended phosphate groups. We have been able to show that the deformation of these beads is uncorrelated with the dynamics of the other beads (including between a sugar and the associated base) or with the bead's own position. However, there is correlation with the bead's orientation, leading to indirect correlations with the motions of other beads. Consequently, in our mesoscopic DNA repre-

sentation, sugar or extended phosphate deformations will become rotations during the fitting process and implicitly affect the dynamics of other beads.

In contrast to the backbone beads, base deformations in models 1 and 2 are too small to have significant effects. With base pair beads (model 4), however, we have seen a moderate correlation (0.25–0.33) with the rigid-body motions of other base pairs. However, in this case we need to question the accuracy of the calculations. The number of variables n describing internal dynamics is three times higher for an AT base pair (57 variables, without hydrogens) than for an extended phosphate group (18 variables), and the number of available conformations N is only 1900 for each trajectory. Increasing the number of variables strongly affects the accuracy of the correlation matrix elements, which is not compensated for by an increase in sampling. Although the uncertainty of the canonical correlation values is difficult to estimate, it certainly does not scale far from $[n/N]^{1/2}$ (Girshick, 1939).

The most appropriate model for subsequent studies will depend on the property to be investigated. It is clear that none of the models proposed can be used for structural determination from NMR or x-ray data, because this requires atomic resolution. It should also be noted that our models do not resolve sugar ring puckering. However, we can specify the generic use of the various mesoscopic representations:

Model 2 (and, a fortiori, model 1) is appropriate for analyzing the dynamic behavior of double-stranded DNA

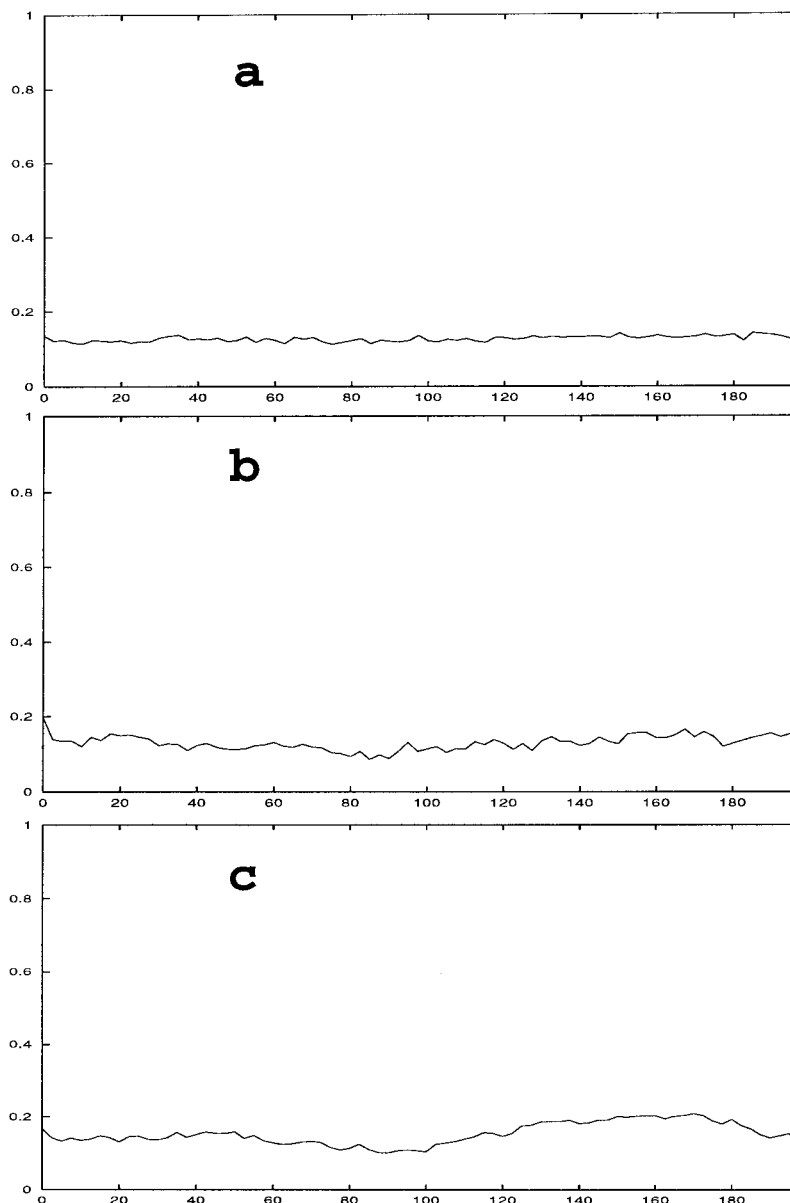


FIGURE 5 Canonical cross-correlation functions between the conformation of the sugar ring of T26 and (a) the conformation of the extended phosphate group of T26, (b) the orientation of the base T26, and (c) the position of the base T26. Shown for simulation 1-B.

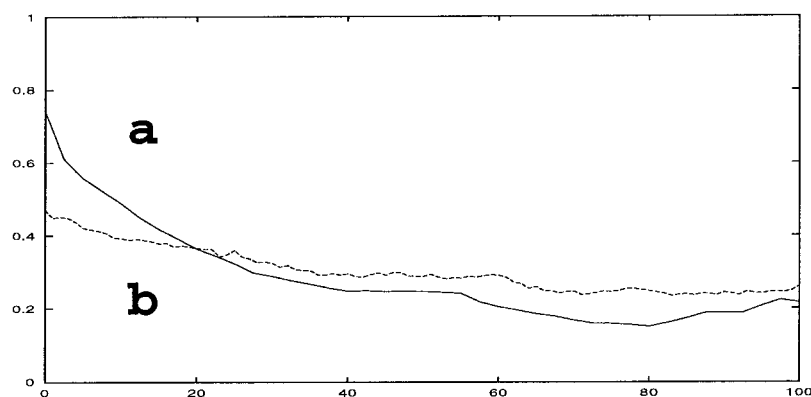
sequences, using the inter- and intrabase or base pair helical parameters defined at the EMBO meeting in Cambridge (Dickerson et al., 1989). These parameters describe the orientation and the position of single bases or of complementary base pairs. This information is conserved in model 2 because we have shown that no correlation exists between the deformation of the extended phosphate or of the sugar puckering conformations and the rigid-body motions of the bases. Thus using three rigid objects per nucleotide in place of an all-atom representation should not cause any loss of accuracy in the determination of the helical parameters. It should be noted that model 2 (or model 1) should be applicable to both single- and triple-stranded helices and that, provided the effective force fields are appropriate, nothing opposes base pair disruption in these models.

Model 4 is more restrictive, as it implies that complementary bases always remain hydrogen bonded. This is a

valid assumption for the simulation of double-stranded DNA sequences when the relative motions of individual bases within a pair can be neglected and could reasonably apply to long DNA fragments at room temperature for periods in the nanosecond range. This model can still be used for studying interbase pair parameters, because the orientation and position of individual base pairs are, at most, very weakly coupled to any other rigid bodies, but its use naturally implies that intra-base pair parameters can no longer be monitored.

Models 3 and 5 correspond to rougher approximations. Although we have not calculated the correlation between the internal deformations of the corresponding rigid bodies and their translation and rotation (because the ratio of the number of configurations to the number of variables is too small), one can expect that as the distance between two rigid bodies increases, this correlation decreases. Therefore it is

FIGURE 6 Canonical cross-correlation functions between (a) translational motions of bases A25 and T26 and (b) rotational motions of bases A11 and A12. Shown for simulation 1-B.



certainly possible to use these models for simplifying the calculation of long-range interactions. This also implies force-field simplifications, such as the use of multipole expansions for treating the electrostatic interactions between distance beads.

It is finally remarked that it is perfectly feasible to combine several of these models in a “reaction center” approach, using either an all-atom representation or the more refined bead models (1, 2, or 4) at the point of interest, combined with more approximate representations (models 3 or 5) and correspondingly simplified force fields (harmonic?) for more distant nucleotides.

The quality of simulations using a bead model depends both on the validity of freezing internal degrees of freedom of the beads and on the treatment of the interbead interactions. A number of authors have used the notion of beads to study long DNA filaments, especially the hydrodynamics of supercoiled DNA (Tan and Harvey, 1989; Chirico and Langowski, 1994; Jian et al., 1998; Klenin et al., 1998). However, most authors used a priori defined beads, making it difficult to predict whether the associated errors can be safely ignored for a specific application (although some of these models correspond to our results, as, for example, in the case of the Tan and Harvey (1989) base plane approach and our model 4). Our study shows how this difficulty can be overcome. In the discussion above we are able to specify where models 1, 2, and 4 can be used safely, with no significant loss of accuracy in the determination of helical parameters due to the introduction of beads. In addition, and

in contrast to nearly all present models, this approach makes it possible to conserve sequence-dependent features.

Interbead potentials remain to be calculated, but several general remarks can again be made. First, a simple, spherically symmetrical, united-atom model for each bead is certainly not appropriate, because it excludes the anisotropic effects due to shape and to charge distribution. Similarly, a refined treatment of the interaction forces and torques between two linked rigid bodies cannot be reduced to harmonic potentials. We return to this point shortly.

By extracting data corresponding to the base pair bead model 4 from our all-atom dynamics, we have calculated the local twisting, bending, and stretching rigidities of DNA. Experimental estimations of rigidities are roughly $1\text{--}4 \times 10^{-19}$ erg·cm (Barkley and Zimm, 1979; Thomas et al., 1980; Millard et al., 1988) for twisting and $2\text{--}3 \times 10^{-19}$ erg·cm (Barkley and Zimm, 1979; Millard et al., 1988) for bending. Our values are in reasonable agreement with these results. In contrast, the stretching rigidity resulting from the experiments of Cluzel et al. (1996) and Smith et al. (1996), ~ 1100 pN, is more than three times smaller than the average value we obtain at the level of individual base pairs (3760 pN). However, it should not be forgotten that the experimental value refers to the overall stretching of long DNA polymers, while our computation refers to an average local value between two successive base pairs. We examined this point by considering more distant base pairs for computing S with Eq. 4c. By using either the distance between the first and the last base pairs analyzed or the sum

TABLE 4 Stretching (S), bending (B), and torsional (C) rigidities of the bead model calculated from model 4

Property	1-B	1-BAB	2-B	2-BAB
l_0 (nm)	0.361	0.390	0.366	0.362
$\langle \Delta l^2 \rangle$ (nm ²)	5.5×10^{-4}	4.5×10^{-4}	5.8×10^{-4}	3.7×10^{-4}
S (pN)	3605 ± 1486	4022 ± 1630	2746 ± 756	4007 ± 456
$\langle \Delta \beta^2 \rangle$ (rad ²)	7.92×10^{-3}	6.73×10^{-3}	6.43×10^{-3}	8.13×10^{-3}
B ($\times 10^{-19}$ erg·cm)	4.5 ± 1.9	5.4 ± 2.0	5.2 ± 1.7	4.9 ± 2.3
$\langle \Delta \gamma^2 \rangle$ (rad ²)	15.00×10^{-3}	7.23×10^{-3}	11.00×10^{-3}	6.79×10^{-3}
C ($\times 10^{-19}$ erg·cm)	1.20 ± 0.55	2.65 ± 0.82	1.53 ± 0.44	2.78 ± 1.0

l_0 is the average (over the configurations and over the beads) distance between the centers of mass of two consecutive base pairs, and $\langle \Delta l^2 \rangle$ is the corresponding RMS fluctuation (nm). $\langle \Delta \beta^2 \rangle$ and $\langle \Delta \gamma^2 \rangle$ are, respectively, the equivalent average RMS fluctuations of the bending and torsion angles between two base pairs.

of the interbead distances (a better estimate of the length of the DNA filament), we obtain a significant decrease in S (1530–2260 pN). If we had studied a longer DNA fragment we could reasonably hope to approach the experimental result. The important point is that this decrease clearly reflects a negative correlation between the stretching of successive interbead distances. This correlation will be totally absent in a priori bead models that are parameterized at the interbead level to reproduce macroscopic elastic properties. The consequences of such approximations, like those related to the arbitrary choice of beads, are again difficult to predict.

We intend to use the rigorous bead definitions set out here for future studies of DNA's elastic and hydrodynamic properties. For this, model 4 should be appropriate and already represents a 6.5 times reduction in the number of variables compared to an all-atom simulation. Force-field development, including sequence effects, will certainly require denser sampling of all-atom trajectories. Long simulations are also important, as shown by the small delayed correlation between bead deformation and the rigid-body motions of other beads observed on the basis of short simulations (Genest, 1996), but they are absent here. The resulting models will be appropriate for energy minimization and Monte Carlo simulations and can be extended to dynamics studies by the addition of viscous solvent effects.

REFERENCES

- Allison, S. A., and J. A. McCammon. 1984. Multistep Brownian dynamics: application to short wormlike chain. *Biopolymers*. 23:363–375.
- Barkley, M. D., and B. Zimm. 1979. Theory of twisting and bending of chain macromolecules: analysis of the fluorescence depolarization of DNA. *J. Chem. Phys.* 70:2991–3007.
- Briki, F., and D. Genest. 1994. Canonical analysis of correlated atomic motions in DNA from molecular dynamics simulation. *Biophys. Chem.* 52:35–43.
- Briki, F., and D. Genest. 1995. Rigid body motions of sub-units in DNA: a correlation analysis of a 200 ps molecular dynamics simulation. *J. Biomol. Struct. Dyn.* 12:1063–1082.
- Cheatham, T. E., III, J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman. 1995. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA and proteins. *J. Am. Chem. Soc.* 117:4193–4194.
- Chirico, G., and J. Langowski. 1994. Kinetics of DNA supercoiling studied by Brownian dynamics simulation. *Biopolymers*. 34:415–433.
- Cluzel, P., A. Lebrun, C. Heller, R. Lavery, J. L. Viovy, D. Chatenay, and F. Caron. 1996. DNA: an extensible molecule. *Science*. 271:792–794.
- Cornell, W. D., P. Cieplak, C. I. Bayley, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmayer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.
- Darden, T. A., D. M. York, and L. G. Pedersen. 1993. Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
- Dickerson, R. E., M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, O. Kennard, R. Lavery, H. C. M. Nelson, W. K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. M. Soumpasis, C. S. Tung, E. von Kitzki, A. H. J. Wang, and V. B. Zhurkin. 1989. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* 205:787–791.
- Feig, M., and B. M. Pettitt. 1998. Structural equilibrium of DNA represented with different force fields. *Biophys. J.* 75:134–149.
- Flatters, D., and R. Lavery. 1998. Sequence-dependent dynamics of TATA-box binding sites. *Biophys. J.* 75:372–381.
- Flatters, D., M. A. Young, D. L. Beveridge, and R. Lavery. 1997. Conformational properties of the TATA-box binding sequence of DNA. *J. Biomol. Struct. Dyn.* 14:757–765.
- Gabb, H. A., C. Prevost, G. Bertucat, C. H. Robert, and R. Lavery. 1997. Collective-variable Monte Carlo simulation of DNA. *J. Comp. Chem.* 18:2001–2011.
- Gaudin, F., G. Lancelot, and D. Genest. 1997. Search for rigid subdomains in DNA from molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 15:357–367.
- Genest, D. 1996. How long does DNA keep the memory of its conformation? A time-dependent canonical correlation analysis of molecular dynamics simulation. *Biopolymers*. 38:389–399.
- Genest, D. 1998. Motion of groups of atoms in DNA studied by molecular dynamics simulation. *Eur. Biophys. J.* 27:283–289.
- Genest, D., and P. Wahl. 1978. Fluorescence anisotropy decay due to rotational Brownian motion of ethidium intercalated in double strand DNA. *Biochim. Biophys. Acta.* 52:502–509.
- Girshick, M. A. 1939. On the sampling theory of roots of determinantal equations. *Ann. Math. Stat.* 10:203–224.
- Guéron, M., M. Kochoyan, and J. L. Leroy. 1987. A single mode of base pair opening drives imino proton exchange. *Nature*. 328:89–92.
- Hery, S., D. Genest, and J. C. Smith. 1998. X-ray diffuse scattering and rigid-body motions in crystalline lysozyme probed by molecular dynamics simulation. *J. Mol. Biol.* 279:303–319.
- Hogan, M. E., and O. Jardetzky. 1980. Internal motions in deoxyribonucleic acid II. *Biochemistry*. 19:3460–3468.
- Jian, H., T. Schlick, and A. Vologodskii. 1998. Internal motion of supercoiled DNA: Brownian dynamics simulations of site juxtaposition. *J. Mol. Biol.* 284:287–296.
- Kim, Y., J. H. Gieger, S. Hahn, and P. B. Sigler. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature*. 365:512–520.
- Klenin, K., H. Merlitz, and J. Langowski. 1998. A Brownian dynamics program for the simulation of linear and circular DNA and other wormlike chain polyelectrolytes. *Biophys. J.* 74:780–788.
- Lafontaine, I., and R. Lavery. 1999. Reduced coordinate models of nucleic acids. *Curr. Opin. Struct. Biol.* (in press).
- Lavery, R. 1994. Modeling nucleic acids: fine structure, flexibility and conformational transitions. *Adv. Comput. Biol.* 1:69–145.
- Leroy, J. L., D. Broseta, and M. Guéron. 1985. Proton exchange and base pair opening of poly(rA)-poly(rU) and poly(rI)-poly(rC). *J. Mol. Biol.* 184:165–168.
- McLachlan, A. D. 1979. Gene duplication in the structural evolution of chymotrypsin. *J. Mol. Biol.* 128:49–79.
- Millard, D. P., K. M. Ho, and M. J. Aroney. 1988. Modification of DNA dynamics by platinum drug binding: a time-dependent fluorescence depolarization study of the interaction of *cis*- and *trans*-diamminedichloroplatinum(II) with DNA. *Biochemistry*. 27:8599–8606.
- Olson, W. K. 1996. Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.* 6:242–256.
- Packer, M. J., and C. A. Hunter. 1998. Sequence dependent DNA structure: the role of the sugar-phosphate backbone. *J. Mol. Biol.* 280:407–420.
- Pearlman, D. A., D. A. Case, G. W. Caldwell, W. S. Ross, T. E. Cheatham, III, D. M. Ferguson, G. L. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman, editors. 1995. AMBER 4.1. University of California at San Francisco, San Francisco, CA.
- Ryckaert, J. P., G. Cicotti, and H. J. C. Berendsen. 1977. Numerical integration of Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comp. Physiol.* 23:327–341.
- Schlick, T. 1995. Modeling superhelical DNA: recent analytical and dynamic approaches. *Curr. Opin. Struct. Biol.* 5:145–262.
- Schlick, T., and W. K. Olson. 1992. Supercoiled DNA energetics and dynamics by computer simulation. *J. Mol. Biol.* 223:1089–1119.
- Smith, S. B., Y. Cui, and C. Bustamante. 1996. Overstretching DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*. 271:795–799.

- Sproun, D., M. A. Young, and D. L. Beveridge. 1998. Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture. Theoretical considerations of the A to and from B transition. *J. Phys. Chem.* 102:4658–4667.
- Tan, R. K., and S. C. Harvey. 1989. Molecular mechanics model of supercoiled DNA. *J. Mol. Biol.* 205:573–591.
- Thomas, J. C., S. A. Allison, C. L. Appelof, and M. Schurr. 1980. Torsion dynamics and depolarization of fluorescence of linear macromolecules. II. Fluorescence polarization anisotropy measurements on a clean viral ϕ 29 DNA. *Biophys. Chem.* 12:177–188.
- van Gunsteren, W. F., and H. J. C. Berendsen. 1986. GROMOS 87 package, Biomos. Biomolecular Software, University of Groningen, Groningen, the Netherlands.
- Wahl, P., C. Paoletti, and J. B. Le Pecq. 1970. Decay of fluorescence emission anisotropy of the ethidium bromide-DNA complex. Evidence for an internal motion in DNA. *Proc. Natl. Acad. Sci. USA.* 65:417–421.
- Zhurkin, V. B., N. B. Ulyanov, A. A. Gorin, and R. L. Jernigan. 1991. Static and statistical bending of DNA evaluated by Monte Carlo simulations. *Proc. Natl. Acad. Sci. USA.* 88:7046–7050.