ELSEVIER

Biophysical Chemistry 55 (1995) 7–29

Biophysical Chemistry

Flexing and folding double helical DNA

Wilma K. Olson *, Marla S. Babcock, Andrey Gorin, Guohua Liu, Nancy L. Marky, Jennifer A. Martino, Scott C. Pedersen, A.R. Srinivasan, Irwin Tobias, Timothy P. Westcott, Peisen Zhang

Department of Chemistry, Rutgers, the State University of New Jersey, New Brunswick, NJ 08903, USA

Abstract

DNA base sequence, once thought to be interesting only as a carrier of the genetic blueprint, is now recognized as playing a structural role in modulating the biological activity of genes. Primary sequences of nucleic acid bases describe real three-dimensional structures with properties reflecting those structures. Moreover, the structures are base sequence dependent with individual residues adopting characteristic spatial forms. As a consequence, the double helix can fold into tertiary arrangements, although the deformation is much more gradual and spread over a larger molecular scale than in proteins. As part of an effort to understand how local structural irregularities are translated at the macromolecular level in DNA and recognized by proteins, a series of calculations probing the structure and properties of the double helix have been performed. By combining several computational techniques, complementary information as well as a series of built-in checks and balances for assessing the significance of the findings are obtained. The known sequence dependent bending, twisting, and translation of simple dimeric fragments have been incorporated into computer models of long open DNAs of varying length and chemical composition as well as in closed double helical circles and loops. The extent to which the double helix can be forced to bend and twist is monitored with newly parameterized base sequence dependent elastic energy potentials based on the observed configurations of adjacent base pairs in the B-DNA crystallographic literature.

Keywords: B-DNA; Base sequence; Computer simulation; Elastic energy; Supercoiling

1. Introduction

DNA is a deceptively simple molecule in terms of its chemical composition and uniform secondary structure. The structural features of the molecule are quite subtle and not obvious from what may seem at first glance to be a uniform helical structure. Small almost imperceptible changes in bending, twisting, and stretching at the level of the four common base pairs introduce perturbations in the familiar double helix, that not only produce structural motifs which facilitate the specific binding of proteins and other ligands but also reveal themselves in a higher order folding of the chain itself. Such folding is essential to the packaging of DNA within the narrow confines of the nucleus as well as an important facet of biological processes that involve recognition and/or long-range interactions [1,2].

The way in which the sequence dependent variability of DNA reveals itself in biological systems depends on the level of structural resolution. At the

* Corresponding author.

SSDI 0301-4622(94)00139-1

base pair level, the sequence dependent features of the double helix modulate groove structure, widening and narrowing the distance between the antiparallel strands of the sugar-phosphate backbone and facilitating the association of water, drugs, proteins, and other molecules with the double helix [3–9]. At polymeric lengths where small deformations of local structure can accumulate over a large number of residues, the overall macroscopic shape is the most important characteristic of the DNA [10,11]. Indeed, it is in chains of about 150 base pairs (bp) or longer that specific sequences appear to exhibit strong natural curvature, running anomalously slowly on gels [12–16] and exhibiting circular shapes under the electron microscope [17,18]. These curved pieces are thought to play a structural role in long naturally occurring DNA. It has been suggested, for example, that intrinsically curved chain fragments might fix the branching of a supercoiled molecule, with the curved pieces clustering in the hairpin loops of the interwound arms [19]. The presence of bound proteins, such as the histones in the nucleosome, similarly influences the shape of a long DNA [20–23].

There is accumulating evidence that base sequence also governs the deformability or the intrinsic 'stiffness' of the double helix. For example, some base pair steps appear to bend preferentially in a particular direction [24]. The bending flexibility is particularly relevant to DNA–protein interactions as many DNA-binding proteins bend the double helix [25–27]. Moreover, some long repeating sequences under specific solution conditions appear to resist folding around proteins, such as the histones in the nucleosome core particle [28–32]. The twisting motions are important determinants of DNA supercoiling. Small changes in local helical twist can lead to large-scale changes in the overall three-dimensional shape of a long closed circular DNA [33,34]. The specific chemical contacts of individual base pairs together with their sequential context (i.e., the chemical identity of the heterocyclic side groups immediately adjacent to each base pair in the chain) presumably determine the intrinsic equilibrium rest states and the range of movements of the double helix.

In this article we focus attention on the local base pair parameters that determine the long-range organization and reveal themselves in the overall macroscopic shape of DNA. Specifically, we concentrate on the local angular parameters between adjacent base pair planes: twist Ω, roll ρ, and tilt τ. The local sequence-dependent translations of adjacent residues also affect the overall shape of the chain [35,36], but are not discussed in this review. Here, for simplicity, we define the chain in terms of a sequence of virtual bonds of fixed length along successive base pair normals, an approximation consistent with the observed lateral displacements of base pairs (i.e., mean shift and slide near zero). In the next section we examine the ranges of local conformations of adjacent base pairs that are revealed in the accumulating DNA X-ray crystallographic literature. We use this information to study effects of base sequence on selected properties of linear polymers. We then describe the methodology needed to study spatially constrained DNA chains. We conclude with several illustrations of the effects of base sequence and/or bound ligands on the preferred shapes of such DNA.

2. Local base pair geometry

The growing numbers of solved nucleic acid crystal structures are today shedding important new clues for deciphering the spatial code hidden within the DNA double helix. Particularly useful in this regard are the orientational and translational perturbations of adjacent base pairs in the solid state structures compared to the perfectly ordered arrangements of these groups in ideal duplex models. There are currently nearly 50 solved B-DNA crystal structures with coordinates deposited in the Nucleic Acid Database [37]. The number of solved structures are sufficient to begin to extract 'rules' describing the local features of DNA structure [24,38].

The distributions of rotational states illustrated in Fig. 1 are based on 426 different dinucleotide steps (213 unique dimers) from 39 B-DNA crystal structures in the database (A. Gorin, unpublished data). Complexes with drugs or proteins and chemically modified duplexes are not included in the survey. Terminal dimer steps, which frequently adopt alternate conformations with positive propeller twist, etc., are also omitted. Such steps are subject to packing effects in some crystal structures [39–41] as well as known to fray and initiate melting of the double

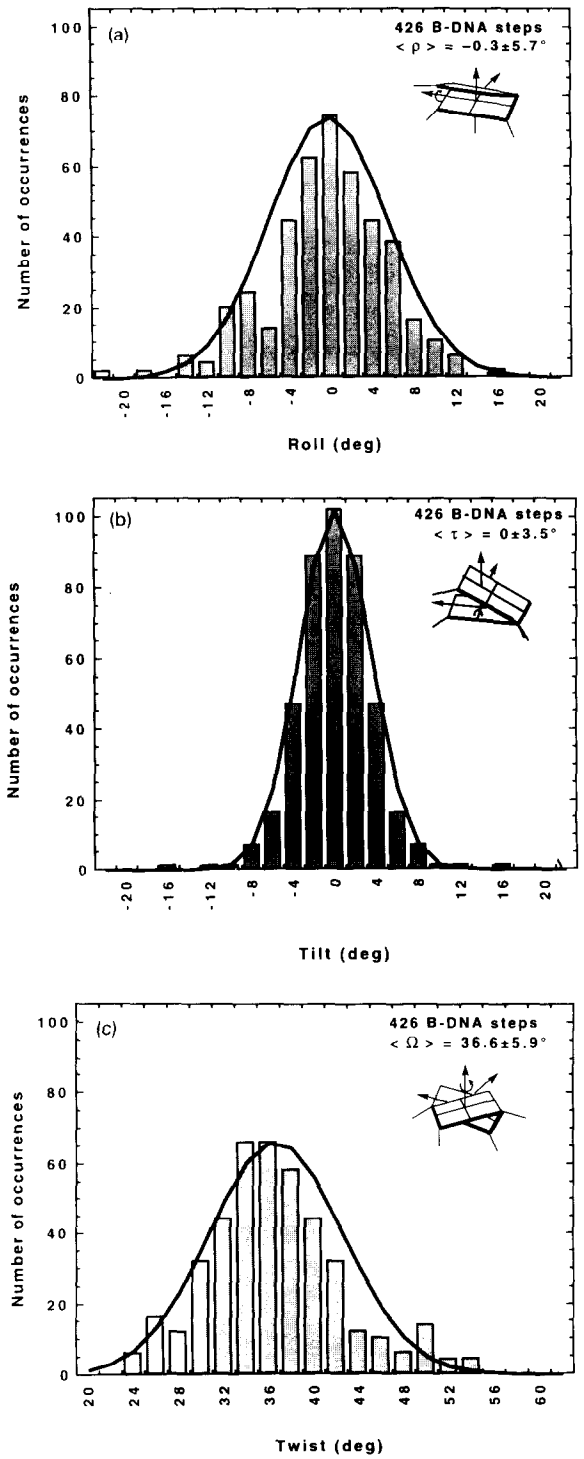

Fig. 1. Distributions, mean values, and standard deviations of: (a) roll, (b) tilt, and (c) twist angles at 426 base pair steps from 39 B-DNA structures in the Nucleic Acid Database [37].

helix in aqueous salt solution [42]. The observed variations in base pair angles in the figure confirm many existing ideas about DNA structure, and also provide some new insights. Significantly, the mean roll and tilt angles are 0° and the average twist angle is ca. 36°, consistent with the orientation of adjacent base pairs in the canonical double helical fiber diffraction model [43]. In other words, the bases are on average perpendicular to the chain axis and the helical repeat is approximately 10 residues per turn. In addition, the mean values and fluctuations of the angles are independent of the method [24,44] used to determine base pair parameters (e.g., $\langle \Omega \rangle = 36.6 \pm 5.9°$, $\langle \rho \rangle = -0.3 \pm 5.7°$, $\langle \tau \rangle = 0 \pm 3.5°$ versus $\langle \Omega \rangle = 35.9 \pm 5.5°$, $\langle \rho \rangle = -0.7 \pm 6.3°$, and $\langle \tau \rangle = -0.3 \pm 3.8°$ using the equations in references [24] and [44], respectively; the former values are reported in Fig. 1). As evident from the widths of the distributions, it is easier to bend the DNA through rolling rather than tilting motions. This observation confirms an original idea of Zhurkin and co-workers [45] based on semi-empirical potential energy calculations and also supports conventional models of DNA bending around a protein surface [46–48].

The widths of the distributions in Fig. 1 provide estimates of the forces opposing the bending and twisting deformations of the double helix. The Gaussian nature of the distributions suggests that the energies associated with fluctuations in individual base pair angles are harmonic. The compensatory geometries of different base pair steps (see below) together with the uniform spread of data provide a reasonable basis for extracting the force constants impeding deformations of random sequence DNA from the angular distributions. Elastic force constants (A_1, A_2, C) of rolling, tilting, and twisting are thus defined in terms of the mean-square fluctuations of base pair angles in the crystallographic sample, the Boltzmann constant k, and the absolute temperature T:

$$A_1 = \frac{kT}{\langle \rho^2 \rangle}, \; A_2 = \frac{kT}{\langle \tau^2 \rangle}, \; C = \frac{kT}{\langle \Omega^2 \rangle} \qquad (1)$$

The mean-square fluctuations in these expressions correspond to the angular deviations that raise the energy per base pair step by 0.5 kT. As a first approximation and despite some preliminary crystal-

lographic evidence to the contrary [24,38], correlations between angular parameters and the corresponding energetic cross terms involving $\langle \rho\tau \rangle$, $\langle \rho\Omega \rangle$, and $\langle \tau\Omega \rangle$ are ignored here.

3. Polymer properties

The force constants defined above can be compared against other measures of bending and twisting extracted by fitting simple elastic rod models to various physical properties of polymeric DNA in dilute solution (e.g., light scattering, chain cyclization kinetics, triplet anisotropy decay, etc.). One such comparison involves the persistence length, a, a measure of the distance over which the direction of the DNA is maintained. Mathematically, a is the mean projection in the limit of infinite chain length of a flexible DNA along its initial direction [49]. If both the rolling and tilting components of bending fluctuate independently and symmetrically about a 0° rest state, as is the case here, the persistence length is given by the simple expression [11],

$$a = \frac{\Delta s}{1 - \langle \cos\rho \rangle \langle \cos\tau \rangle} \tag{2}$$

where Δs is the spacing (i.e., virtual bond length) between adjacent base pairs and the angle brackets

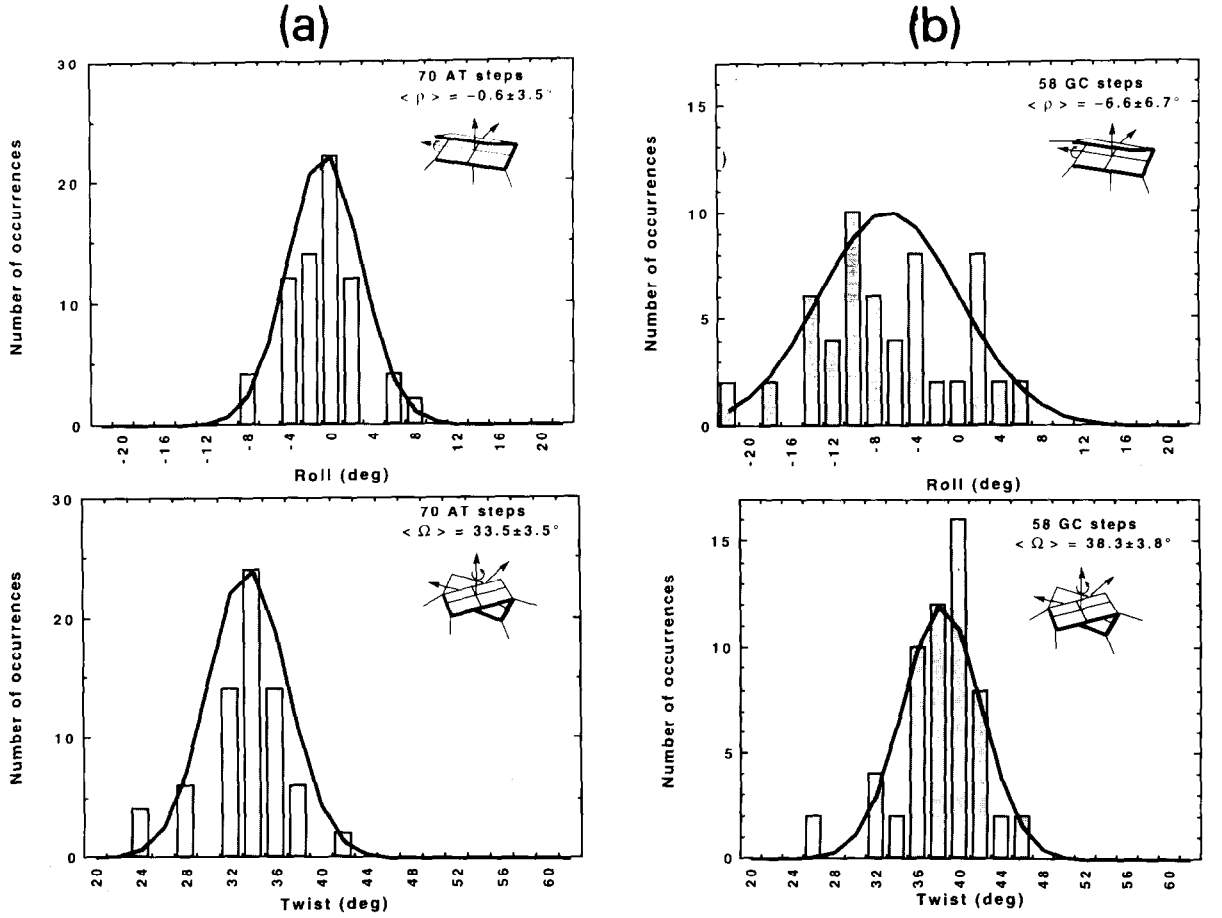

Fig. 2. Distributions, mean values, and standard deviations of the roll and twist angles at: (a) 70 AT and (b) 58 GC base pair steps in B-DNA crystal structures in the Nucleic Acid Database [37].

denote Boltzmann averages. Because the distributions of roll and tilt are Gaussian, the average cosines in Eq. 2 can be further simplified using standard integrals [50] so that the expression for the persistence length reduces to:

$$a = \frac{\Delta s}{1 - \exp\left[\frac{-\langle \Delta\rho^2 \rangle}{2}\right]\exp\left[\frac{-\langle \Delta\tau^2 \rangle}{2}\right]} \qquad (3)$$

Substitution of the numerical values listed in Fig. 1, expressed in radians, into this equation yields a persistence length of 501 Å, a value in remarkable agreement with the measured extension of mixed sequence DNA in aqueous salt solutions (10^{-2} M or higher concentration Na^+) [51–53]. A comparable limiting value (419 Å), also within the limits of experimental observation, is obtained when the base pair angles in the same B-DNA sequences are defined by the mathematically rigorous equations in reference [44]. Surprisingly, the range of B-DNA conformations observed in the ordered crystalline environment is equivalent to the conformational flexibility of the polymer in dilute solution.

We also note from Fig. 1 that the net fluctuations in roll are comparable to those in twist. The overall bending stiffness as measured by the inverse of the mean fluctuations in the two bending components,

$$A = \left[\frac{2kT}{\langle \Delta\rho^2 \rangle + \langle \Delta\tau^2 \rangle}\right] \qquad (4)$$

is greater than the torsional stiffness C (Eq. 1). The above definition of A corresponds in the limit of isotropic bending (i.e., $\langle \rho^2 \rangle = \langle \tau^2 \rangle$) to the classical model of a worm-like chain [52]. The ratio, A/C, describing the ease of overall bending versus twisting adjacent base pairs, is thus greater than unity, a result in agreement with the measured tendencies of bending compared to twisting deduced from chain cyclization kinetics experiments [55,56] but counter to recent interpretations of time-resolved fluorescence polarization and dynamic light scattering measurements of DNA plasmids [54].

4. Base sequence dependent structure

Sufficient crystallographic data are beginning to accumulate so that it is becoming possible to extract sequence-dependent 'rules' of twisting and bending successive base pairs along the double helix. While not all of the common dimer steps are well represented in the current X-ray literature, there are enough examples of the self-complementary purine–pyrimidine dimers, AT and GC (i.e., $(ApT)_2$ vs. $(GpC)_2$), to see significant differences between them. Fig. 2 reports the distributions of roll and twist angles in these two dimers (70 and 58 observations, respectively), taken from a recent analysis [24] of B-DNA structures in the Nucleic Acid Database. The collected data show dramatic differences in both mean angular values and root-mean-square fluctuations with modification of chemical sequence. The AT step is essentially planar on average ($\langle \rho_{AT} \rangle = -0.6°$), whereas the GC step is decidedly bent into the major groove ($\langle \rho_{GC} \rangle = -6.6°$). The AT dimer is also noticeably stiffer in terms of the smaller breadth of observed roll angles compared to the GC dimer ($\langle \Delta\rho^2_{AT} \rangle^{1/2} = 3.5°$ vs. $\langle \Delta\rho^2_{GC} \rangle^{1/2} = 6.7°$). Furthermore, because the roll and twist angles vary in a correlated fashion, the AT steps have a lower mean twist than the GC steps ($\langle \Omega_{AT} \rangle = 33.5°$ vs. $\langle \Omega_{GC} \rangle = 38.3°$). The fluctuations in twist, however, are comparable in the two dimers ($\langle \Delta\Omega^2_{AT} \rangle^{1/2} = 3.5°$ vs. $\langle \Delta\Omega^2_{GC} \rangle^{1/2} = 3.8°$). Other base pair steps, for which there are fewer experimental examples, appear to counterbalance the geometries of these two dimers, thereby accounting for the mean parameters of all DNA base pairs. For example, the positive roll and low twist of CG and GG steps could compensate for GC wedges of the opposite sense, while the high twist of TA may offset the apparent underwinding of AT [24].

The implications of such differences in local bending and twisting can be far reaching, particularly when the equilibrium structures of specific dimer steps are regularly repeated in a long polymer chain. A small roll angle, for example, inclines base pairs with respect to the double helical axis. Introduction of such perturbations over a half helical turn of DNA produces a wedge in the double helix, that when propagated in a longer chain leads to a naturally curved static structure [57,58]. The planar curved DNA illustrated in Fig. 3 contains such features. The chain is made up of two kinds of base pair steps, one where successive residues are perfectly planar and twisted by 35.8° and the other where base pairs are

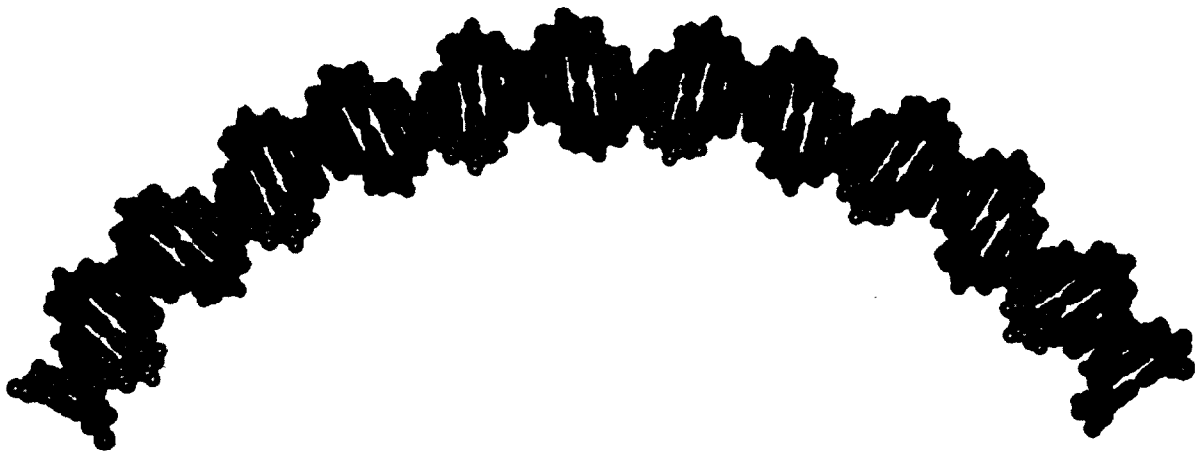

Fig. 3. Computer generated static representation at the all-atom level of a curved 60 bp fragment of DNA built from two kinds of base pair steps: (a) planar steps ($\rho_o = \tau_o = 0°$) that are twisted by $\Omega_o = 35.8°$ between one or more red and blue residues (six such steps per decamer repeat); (b) bent steps with equilibrium rest angles $\rho_o = 6.5°$, $\tau_o = 0°$, $\Omega_o = 36°$ linking consecutive green and yellow base pairs (four such steps per decamer repeat). The distance between the centers of the base pair planes is fixed at 3.4 Å.

assigned a roll angle of 6.5° and a twist angle of 36°. The tilt is fixed at 0° in both kinds of steps and the distance between the centers of base pair planes is 3.4 Å, the latter value being equal to the mean rise of successive base pairs in the B-DNA crystallographic sample. The repeating unit of the curved model is made up of six planar steps (linking one or more red and blue residues) followed by four bent arrangements (between consecutive green and yellow base pairs) and extended over a total of 60 base pairs. The sequential pattern produces a 20°-wedge per decamer repeat and the chain length gives rise to a curve that spans one third the circumference of a circle (i.e., six decamer repeats $\times 20°$ per repeat $= 120°$ arc) [11].

If, in addition, a variation in twist is superimposed upon the sequence-dependent values of the roll angle, the sequence employed in Fig. 3 will follow a superhelical pathway [11]. For example, if the twist of the planar states in this example is reduced from 35.8° to 34.3° (a value typical of the twist in solution of naturally occurring mixed sequence DNA [34,59]), the 60-mer will span 36% of a turn of a right-handed superhelix with a rise per base pair of 1.4 Å. If the twist angle is increased in magnitude, the same sequence will follow a left-handed trajectory.

5. Influence of fluctuations

The extent to which the double helix appears curved in solution depends not only on the equilibrium rest state but also on 'width' of the roll, tilt, and twist fluctuations. In other words, the overall properties of the DNA also reflect the magnitude of the force constants constraining base pair parameters to local minima. As a result, a static equilibrium model can sometimes be quite misleading. For example, a mean roll angle of zero at all base pair steps does not necessarily mean that a DNA is straight. Chains made up of base pair steps that bend in an asymmetric fashion about a planar rest state can exhibit average properties characteristic of an intrinsically curved polymer [11]. Consider adjacent base pairs that tend to roll preferentially into the major rather than the minor groove with energy defined by an 'asymmetric parabola' of the form:

$$E_{skew}(\rho) = \begin{cases} \frac{kT}{2}\frac{(\rho-\rho_o)^2}{\Delta\rho_+^2} & \rho \geq \rho_o \\ \frac{kT}{2}\frac{(\rho-\rho_o)^2}{\Delta\rho_-^2} & \rho < \rho_o \end{cases} \quad (5)$$

Residues obeying this asymmetric function are virtually indistinguishable from dimers described by a symmetric potential of the same form as Eq. 1,

$$E_{sym}(\rho) = \frac{kT}{2} \frac{(\rho - \langle \rho \rangle)^2}{\langle \Delta\rho^2 \rangle} \quad (6)$$

and having the same mean roll angle $\langle \rho \rangle$ and mean-square roll fluctuation $\langle \Delta\rho^2 \rangle$ [11]. The mean roll angle and the root-mean-square fluctuation for dimers obeying the asymmetric potential are, respectively,

$$\langle \rho \rangle = \sqrt{\frac{2}{\pi}} (\Delta\rho_+ - \Delta\rho_-) + \rho_0 \quad (7)$$

and

$$\langle \Delta\rho^2 \rangle^{1/2} = \sqrt{\Delta\rho_+ \Delta\rho_- + \left(1 - \frac{2}{\pi}\right)(\Delta\rho_+ - \Delta\rho_-)^2} \quad (8)$$

Hence, positive roll fluctuations of $\Delta\rho_+ = 2.0°$ and negative roll fluctuations $\Delta\rho_- = 10.2°$ in a dimer with minimum energy at $\rho_0 = 0°$ lead to the same mean positioning of adjacent base pairs as a step with mean roll angle, $\langle \rho \rangle = -6.5°$, and root-mean-square fluctuation, $\langle \Delta\rho^2 \rangle^{1/2} = 6.7°$. Moreover, when incorporated in a repeating DNA polymer, the configuration-dependent properties of the chain as a whole are also equivalent.

The similarity of macromolecular properties in these systems is illustrated in Fig. 4, where the root-mean-square end-to-end distance $\langle r^2 \rangle^{1/2}$ and the mean cosine $\langle \gamma \rangle$ between the normals of terminal base pairs are plotted as a function of chain length N for both an intrinsically curved ($\langle \rho \rangle \neq 0$) DNA with symmetric roll fluctuations and a naturally straight ($\rho_0 = 0$) DNA with asymmetric variations in the roll coordinate. The values are computed by incorporating the average sines and cosines of roll, tilt, and twist angles based on the assumed energy variations in polymer generating matrices that (when combined in order of chain sequence) accumulate exact configuration-dependent average quantities [11,35,60,61]. The fluctuations in twist are set at 5.9° and those in tilt at 3.5° in these examples. The chain repeating sequence is similar to that used to generate the curved static model in Fig. 3. Here the

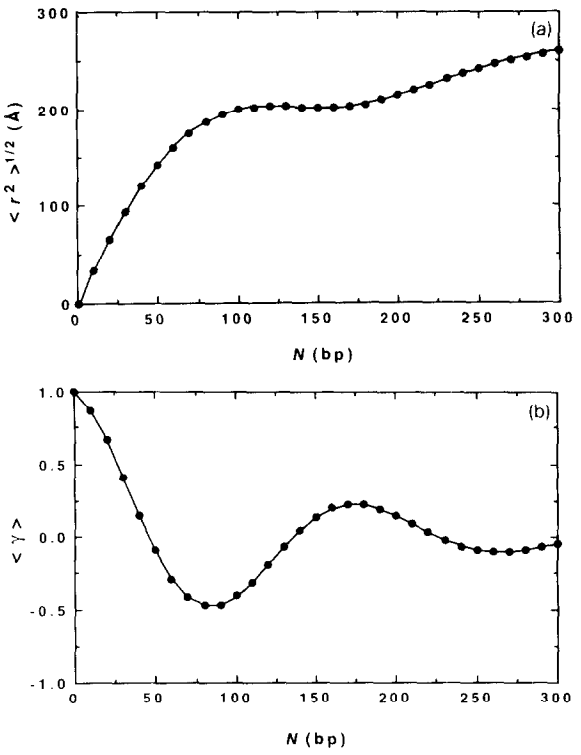

Fig. 4. (a) Unperturbed root-mean-square end-to-end distance, $\langle r^2 \rangle^{1/2}$, and (b) mean cosine between chain ends, $\langle \gamma \rangle$, plotted as a function of chain length N for an intrinsically curved ($\langle \rho \rangle \neq 0$) DNA with symmetric fluctuations in the roll coordinate and a naturally straight ($\rho_0 = 0$) DNA with asymmetric variations in the roll. Parameters describing the profiles of roll energy are as follows: symmetric profile (scattered solid circles) with $\langle \Delta\rho^2 \rangle^{1/2} = 6.7°$ and $\rho_0 = \langle \rho \rangle = -6.5°$; asymmetric profile (solid lines) with $\rho_0 = 0°$, $\Delta\rho_- = 10.2°$, $\Delta\rho_+ = 2.0$. Fluctuations in tilt and roll are assumed to be independent of sequence with $\Delta\tau = 3.5°$, $\Delta\Omega = 5.9°$. Equilibrium rest angles are set at $\tau_0 = 0°$, $\Omega_0 = 36°$. The 10 bp repeating sequence is identical to that in Fig. 3. The excellent agreement between the two profiles shows that the two schemes of bending (symmetric and skewed) are virtually indistinguishable (see text).

polymer is made up of six naturally straight steps followed by four locally perturbed steps, all of which are subject to fluctuations. The straight steps are assigned a mean roll of 0° with root-mean-square fluctuations of 5.7°, while the skewed steps are constructed in two ways: (1) $\langle \rho \rangle = -6.5°$, $\langle \Delta\rho^2 \rangle^{1/2} = 6.7°$; and (2) $\rho_0 = 0°$, $\Delta\rho_- = 10.2°$, $\Delta\rho_+ = 2.0°$). As evident from Fig. 4, chains constructed with both schemes are practically indistinguishable at the macromolecular level. The root-

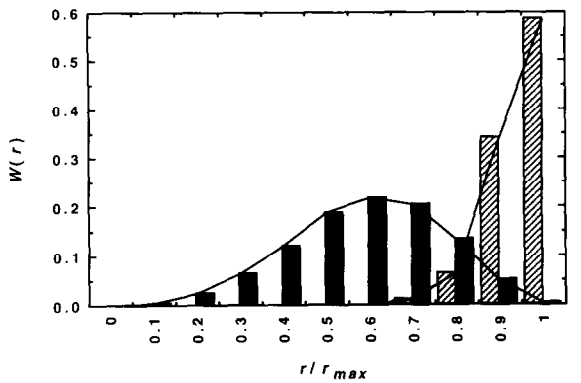

Fig. 5. Histograms of the distributions of end-to-end distances $W(r)$ as a function of relative chain extension, r/r_{max}, obtained from Monte Carlo samples of 100 bp curved (black columns) and straight (striped columns) B-DNA chains. The curved polymer chain is formed by tenfold repetition of six symmetric dimer steps ($\Omega_0 = 36°$, $\rho_0 = 0°$, $\tau_0 = 0°$, $\langle \Omega^2 \rangle^{1/2} = 0°$, $\langle \rho^2 \rangle^{1/2} = 5.7°$, $\langle \tau^2 \rangle^{1/2} = 3.5°$) and four skewed steps ($\rho_0 = 0°$, $\Delta\rho_- = 10.2°$, $\Delta\rho_+ = 2.0°$). The straight chain is made up entirely of symmetric steps. Maximum chain extension of the B-DNA helix, $r_{max} = 340$ Å.

mean-square end-to-end distances of the two types of chains differ by no more than 0.6 Å and the mean cosines by less than 0.004 for repeating sequences up to 300 bp in length. As noted above, the average roll angles of the two chains are identical under these circumstances; the equality further holds for the trigonometric averages ($\langle \cos\rho \rangle$ and $\langle \sin\rho \rangle$) used in the calculations (data not shown). The data associated with the skewed potentials are represented by lines and the values computed for the symmetric potentials with scattered solid symbols in the figure. The good fit of the data is apparent from the nearly perfect overlap of the scattered shapes with the corresponding points on the line.

The importance of fluctuations is further evident from the distributions of end-to-end distances in Fig. 5 compiled from Monte Carlo sampling of 100 bp DNA fragments [62,63]. Individual chain configurations are generated residue-by-residue on the basis of a random number determination of local base pair conformation and weighted according to the Boltzmann factors of their energies. The distributions are obtained by counting the number of chains in a sample of 10^5 configurations within the specified ranges of distances and plotting the numbers with respect to the maximum chain extension of the B-DNA helix, $r_{max} = 3.4\ N$ Å $= 340$ Å. The sample size is based on a comparison of the Monte Carlo average distances and angles with exact values like those in Eq. 2 [11]. As evident from the figure, the distributions are sensitive to chemical sequence (i.e., the equilibrium rest states and fluctuations of individual dimers), particularly at polymeric chain lengths where small deformations of local structure can accumulate over a large number of residues. The range of end-to-end distances of the flexible DNA sequence described above (with six dimer steps fluctuating symmetrically about a planar rest state followed in each repeating unit by four steps with asymmetric roll variation) is compared in the figure against that of an intrinsically straight homopolymeric chain of the same length with symmetric fluctuations in roll and tilt leading to a limiting persistence length of 501 Å. The equilibrium twist, roll, and tilt ($\Omega_0 = 36°$, $\rho_0 = 0°$, $\tau_0 = 0°$, respectively) and the fluctuations in tilt and twist ($\langle \tau^2 \rangle^{1/2} = 3.5°$ and $\langle \Omega^2 \rangle^{1/2} = 0$) are identical in all dimer steps of the two DNAs. (As above, the roll fluctuation $\langle \rho^2 \rangle^{1/2}$ is assigned a value of 5.7° at symmetric dimer steps, while $\Delta\rho_-$ is set at 10.2° and $\Delta\rho_+$ at 2.0° at asymmetric steps.) The conformational flexibility at individual base pair steps allows some combinations of local states to form macromolecular arrangements with ends in close contact. The probability of the chain ends meeting, however, is much greater in the DNA containing dimer steps with asymmetric bending tendencies than in the homopolymer. The intrinsically straight chain with symmetric fluctuations, by contrast, is much more likely to be fully extended. The distribution of end-to-end distances in the skewed chain is essentially identical to that of an intrinsically curved DNA with symmetric roll fluctuations (i.e., the corresponding polymer in which the asymmetric roll fluctuations about a planar rest state are replaced by symmetric fluctuations about an intrinsically bent state, specifically $\rho_0 = -6.5°$ and $\Delta\rho_- = \Delta\rho_+ = 6.7°$). In the skewed DNA fragment the root-mean-square extension, $\langle r^2 \rangle^{1/2}$, is 191 Å and the mean cosine between chain ends, $\langle \gamma \rangle$, is -0.53, whereas in the homopolymer $\langle r^2 \rangle^{1/2} = 302$ Å and $\langle \gamma \rangle = 0.51$. These averages, accumulated in the Monte Carlo sampling, agree within 1% and 3% of the respective exact average distances and orientations obtained using

direct methods [11,35]. In neither case does a single average configuration represent all the DNA structures and account for the properties associated with the whole ensemble of configurations. The distributions become broader and even further from a representative mean configuration as chain length increases [63].

6. Spatially constrained DNA

Compared to their effects in unconstrained linear DNA, the conformational irregularities of successive base pairs play an even more complicated role in spatially constrained molecules. The global folding of a closed circular duplex or a DNA loop clamped at its ends by a protein is tied not only to local bending and twisting preferences but also to the linking number, Lk, or the number of times the two strands of the double helix are wrapped around one another [64]. The latter parameter, an integer, is a topological invariant in the sense that it cannot be changed except by severing (and resealing) one or both of the strands that comprise the DNA plasmid. Bending induced by association with proteins or other ligands may also alter local duplex structure and thereby play an important role in determining the tertiary structure of the closed DNA to which the proteins are bound.

The treatment of closed chain molecules (or, alternatively, open chains with ends confined to a fixed end-to-end separation and orientation) is a long standing problem in polymer physical chemistry that can be attacked from several points of view. Recent work from this laboratory has shown that many of the problems associated with the simulation of spatially constrained DNA can be overcome by taking advantage of curve fitting and finite element methods commonly used in computer-aided design and engineering [65–69]. The curve fitting techniques automatically satisfy the spatial limitations placed on the ends of the DNA while describing the complex curves traced by the double helical axis with a relatively limited number of independent variables. These methods have proved most effective when used in combination with an isotropic elastic rod model of the double helix. In this approximation, where there are no directional preferences on chain bending (i.e., $A_1 = A_2$ in Eq. 1) and the DNA is naturally straight, the total elastic energy (bending and twisting) can be described simply in terms of geometric parameters determined by the helical axis (see below). This simple model cannot incorporate local variations of helical twist or sequence-dependent fluctuations of adjacent residues. It can, however, provide some approximation of the effects of local chain bending on the global folding of DNA by restricting parts of the helical axis to a rigid curved pathway. The latter model is a useful starting point for understanding the effects of bound proteins on overall DNA shape.

The treatment of 'real' DNA chains requires the description of three intertwined closed curves — the double helical axis and two intertwined 'backbone' ribbons. The central axis and one ribbon are required to compute the local twist, while the second ribbon is needed to define the base pair plane. Vectors connecting corresponding points on the helical axis and the two strands are related to the positions of the long and short axes of DNA base pairs and hence to the twisting, rolling, and tilting of adjacent residues. Chains with sequence-dependent elastic character can thus be treated with this formalism. While it is possible to use curve fitting techniques to define the chain axis and backbone ribbons, we have found it more convenient to identify the equilibrium configurations of a constrained duplex by application of a variational principle, which balances the internal elastic energy and the external forces on the DNA (see below).

7. Fourier series representations

A finite three-dimensional Fourier series offers a convenient way to represent the configuration of a DNA chain with the ends constrained to a fixed position and orientation [69]. An expression of the form,

$$\Delta \mathbf{r}(u) = \sum_{m=1}^{M} \{\boldsymbol{a}_m \sin(2\pi m) + \boldsymbol{b}_m(\cos(2\pi u) - 1)\} + \boldsymbol{a}_{M+1} \sin(2\pi(M+1)u) \quad (9)$$

corresponds to the difference in configuration between an arbitrary chain trajectory and an assumed initial pathway of the helical axis. The vectorial coefficients $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ are optimized during the

course of computer simulation with the contour parameter u ranging between zero and unity. The choice of sines and cosines as basis functions assures that the ends of the DNA coincide with those of the starting state (i.e., $\Delta\mathbf{r}(0) = \Delta\mathbf{r}(1) = \mathbf{0}$) while the definition of the coefficient a_{M+1},

$$a_{M+1} = \frac{-1}{M+1} \sum_{m=1}^{M} m a_m \tag{10}$$

further guarantees that the tangents at the ends of the DNA coincide with those of the starting state (i.e., $\Delta\mathbf{r}'(0) = \Delta\mathbf{r}'(1) = \mathbf{0}$). The number of Fourier coefficients is determined by the degree of complexity of chain folding. A greater number of variables is required to describe more highly supercoiled configurations.

A convenient starting configuration of the DNA can be obtained with a Bézier curve of the form:

$$\mathbf{r}_{\text{loop}}(\nu) = \sum_{j=0}^{n} B_j^n(\nu)\mathbf{p}_j \tag{11}$$

The $\mathbf{p}_j$ in this expression are controlling points in Cartesian space that determine the contour of the curve, while the $B_j^n(\nu)$ are coefficients defined in terms of the parameter ν, $0 \le \nu \le 1$, and the number of controlling points, $n + 1$ [70]:

$$B_j^n(\nu) = \frac{n!}{j!(n-j)!} \nu^j (1-\nu)^{n-j} \tag{12}$$

The controlling points $\mathbf{p}_j$ are directly related to the coordinates and tangents at the boundaries of the DNA:

$$\begin{aligned} \mathbf{r}(0) &= \mathbf{p}_0 \\ \mathbf{r}'(0) &= n(\mathbf{p}_1 - \mathbf{p}_0) \\ \mathbf{r}(1) &= \mathbf{p}_n \\ \mathbf{r}'(1) &= n(\mathbf{p}_n - \mathbf{p}_{n-1}) \end{aligned} \tag{13}$$

This property of the curve makes it possible to match the desired constraints on the DNA with an appropriate choice of controlling points. Points $\mathbf{p}_0$ and $\mathbf{p}_n$ coincide with the ends of the chain, while points $\mathbf{p}_0$

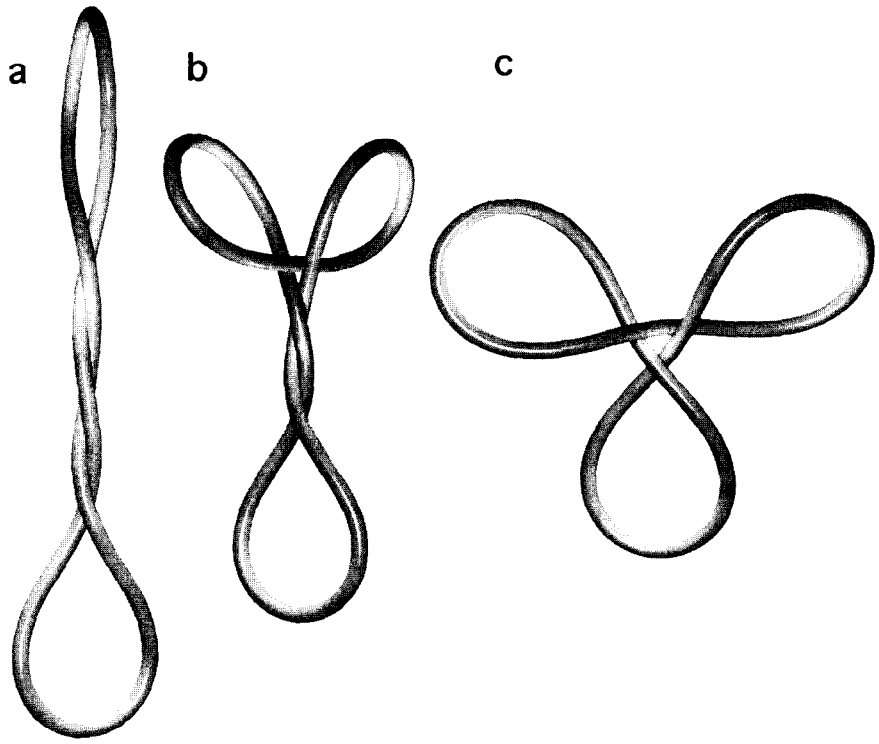

Fig. 6. Molecular representations of selected energy minima of a 1000 bp supercoiled DNA with a linking number difference, $\Delta Lk = 4.0$: (a) linear interwound configuration with $Wr = 3.4$, $E = 6.1$ kT; (b) asymmetric three-lobed branched state with $Wr = 3.3$, $E = 17.6$ kT; (c) symmetric threefold open structure with $Wr = 2.7$, $E = 19.0$ kT.

and $\mathbf{p}_1$ determine the tangent at the start of the DNA, and points $\mathbf{p}_{n-1}$ and $\mathbf{p}_n$ the tangent at the end of the chain. The intervening controlling points, $\mathbf{p}_2$ to $\mathbf{p}_{n-2}$, determine the contour length.

8. DNA as an ideal elastic rod

According to elasticity theory, the local twist is uniform along a normally straight rod with circular cross-section. For such a chain, the total elastic energy (isotropic bending plus twisting) reduces to the simple expression [71],

$$E = \frac{A}{2}\oint \kappa(s)^2 \, ds + \frac{2\pi^2 C}{L}(\Delta Lk - Wr)^2 \qquad (14)$$

where L is the total contour length of the DNA and $\Delta Lk = Lk - Lk_0$ is the linking number difference with respect to the natural linking number of the chain. The curvature κ, a parameter connected to the net bending of base pairs, and the writhing number Wr, a measure of overall chain folding, appearing in the above expression can be evaluated from the curve of the helix axis using standard differential geometric formulas [65,66,72]. Because the local twisting energy is related to the overall shape of the curve in this approximation, there is no need to locate the positions of individual base pairs. The known preference of DNA base pairs to bend into the major and/or minor grooves (i.e., roll preferentially over tilt) and the extensibility of the chain are ignored in the model.

Typical minimum energy states of a 1000 bp closed circular DNA obeying the isotropic rod approximation are shown in Fig. 6. The superhelical density of these configurations, identified using a truncated Newton minimizer and seven Fourier coefficients [73], is set at natural levels ($\Delta Lk/Lk_0 = 0.04$). The bending constant is assigned a value corresponding to a per residue bending fluctuation of 6.3° at 298 K (equivalent to a persistence length of 564 Å), and the ratio, A/C, is taken as unity. A van der Waals' potential with minimum energy at 30 Å separation [66] is introduced to prevent chain self-intersection. The polymer configuration of lowest energy (6.1 kT) is a linear interwound form with end loops roughly perpendicular to one another and a writhing number of 3.4 (Fig. 6a). A secondary minimum of significantly higher total energy (17.6 kT) but nearly equivalent writhing number ($Wr = 3.3$) is an asymmetric three-lobed branched arrangement with chain intertwining in the elongated loop (Fig. 6b). The latter minimum resembles a distorted interwound form with a large folded end loop. The two configurations, while distinctly different in overall shape, have roughly the same number of intramolecular contacts. Because the writhing numbers are of similar magnitude, the twisting energies of the two forms are nearly identical (1.2 kT and 1.5 kT, respectively). As should be clear from these examples, the writhing number is not a unique measure of chain folding, but simply a numerical assessment of the close contacts in the chain. The bending of the DNA axis is much less severe in the linear configuration than in the branched minimum, accounting for the large difference in total energy. The difference in energy would be much smaller in a longer DNA of the same shapes since the bending energy, here the dominant contribution to the total energy, scales inversely with chain length. For example, the bending energy of the three-lobed minimum in Fig. 6b would lie within kT of that of the linear interwound form in a 11400 bp DNA. The occurrence of low-energy branched interwound minima is consistent with the highly branched arrangements of long supercoiled DNA observed in electron microscopy studies [74]. The threefold symmetric open lobed form shown in Fig. 6c is even higher in energy (19.0 kT) than the asymmetric branched minimum, owing primarily to its lower writhing number ($Wr = 2.7$) and consequently higher twisting energy. This state would also be energetically favored in a longer chain.

9. Protein-bound DNA

The binding of a protein to double helical DNA places constraints on the position and orientation of parts of the chain. Proteins may also deform DNA into highly curved structures resembling those associated with specific base sequences [75,76]. The examples described below correspond to a closed circular chain containing a rigid segment wrapped in a left-handed superhelical path around a 'phantom'

protein core and a free connecting loop. The loop has at least one single-stranded scission so that it can assume a torsionally relaxed state. The shape of the loop at equilibrium is simply that of minimum bending energy. Each such equilibrium state has a writhing number which depends on both the shape of the loop and the assumed shape of the rigid, protein-bound segment. If the nicked strand is sealed, the DNA will be characterized by a linking number difference numerically equal to the writhing number. Upon release of the protein, the molecule will assume a new shape, consistent with the value of ΔLk so determined. At this point ΔTw is no longer zero, but equal to $\Delta Lk - Wr$, where Wr is the writhing number of the equilibrium protein-free state.

Fig. 7 shows the variation of molecular shape, as measured by Wr, versus the number of superhelical turns T of a 500 bp torsionally relaxed DNA around a rigid protein core. The writhing numbers are those of the global energy minima identified using Monte Carlo search techniques, an initial Bézier curve with five controlling points, and five Fourier coefficients [69]. The assumed pitch (30 Å) and radius (45 Å) of the bound fragment resemble those of the DNA

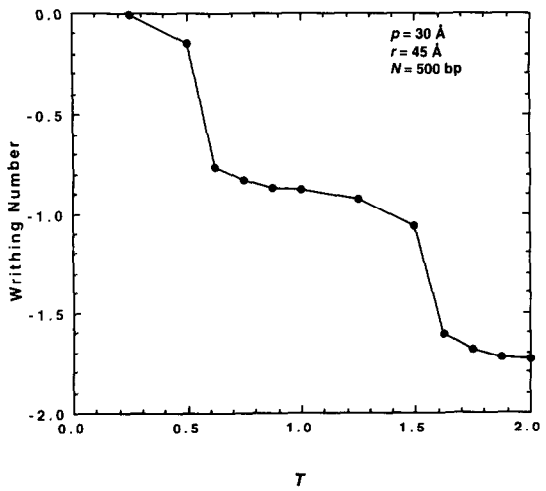

Fig. 7. Variation of the writhing number at the global energy minimum versus the wrapping number T of a torsionally relaxed 500 bp DNA constrained in part around an ideal cylindrical protein core. The fraction of the DNA on the protein core is varied over a range of values between 0–2 superhelical turns. Superhelical radius $r = 45$ Å; pitch $p = 30$ Å. Refer to text for additional details.

wrapped around the histone proteins in the nucleosome core particle [77,78]. Because the chain is nicked, only the bending contribution is included in the elastic energy with the bending constant assigned a value corresponding to a net root-mean-square fluctuation of 5.8° per base pair at 298 K (i.e., $a = 656$ Å). A hard-sphere term with a contact limit of 20 Å is introduced in the calculations to prevent chain self-intersection. The overall decrease in Wr with number of superhelical turns in the minimum energy states is partially expected from the manipulation of physical models. The number of apparent crossings of the unbound loop viewed down the axis of the bound superhelical fragment, is increased as the DNA wrapping is constrained to longer superhelical pathways. Of special note in Fig. 7 are the sharp transitions in Wr with increase in T at approximately half integral turns of the superhelix and the near constancy of Wr vs. T at other values. At low values of superhelical wrapping, the presence of protein has essentially no effect on the writhing number, the value being zero up to $T \approx 0.5$ turns. Over the range $T = 0.5$ to 1.5 turns, Wr is approximately constant at -1, and beyond $T = 1.5$ turns the writhing number is roughly -1.7.

As evident from the structures illustrated in Fig. 8, the transition associated with incremental wrapping around a protein core is a sharp configurational jump (in the sense that a small change in a local parameter produces a large-scale change in three-dimensional structure). Shown here are the global equilibria of a 500 bp DNA with a segment wrapped either (a) $1\frac{1}{2}$ or (b) $1\frac{5}{8}$ superhelical turns around the protein core detailed above. In addition to the sudden collapse of structure that accompanies small incremental wrapping of the DNA (ca. 10 bp in this example), the protein binding core moves suddenly from the inside to the outside of the free DNA segment. The ends of the protein-bound DNA are roughly antiparallel to one another at half turns of the superhelix. The configurational change reflects a critical elastic transition where the imposed bending on the ends of the unbound DNA becomes too severe and the free loop suddenly snaps into a crossed arrangement. Recent extension of the theory of elastic rods shows that the equilibrium configuration of a constrained loop is a function of the distance and angle between terminal residues together

with the total length of the fragment [79]. The crossed arrangement is predicted only if the contour length of the loop is roughly an order of magnitude larger than the distance between its ends and the angle between unit normals of terminal base pairs is greater than ca. 90°.

The equilibrium structures of torsionally relaxed DNA with several bound proteins, such as a minichromosome, can be understood on the basis of the above theoretical principles. For example, no crossings are predicted in a constrained loop unless the through-space separation of chain ends is small compared to the intervening contour length. As the ends of the loop are brought closer together, self-contacts are expected. The examples in Fig. 9 are optimized states of torsionally relaxed isotropic elastic chains of 518 and 1216 bp identified in Monte Carlo studies (J. Martino, unpublished data). The two bound segments of the DNA (each ca. 140 bp in length) are wrapped $1\frac{3}{4}$ turns about the protein core

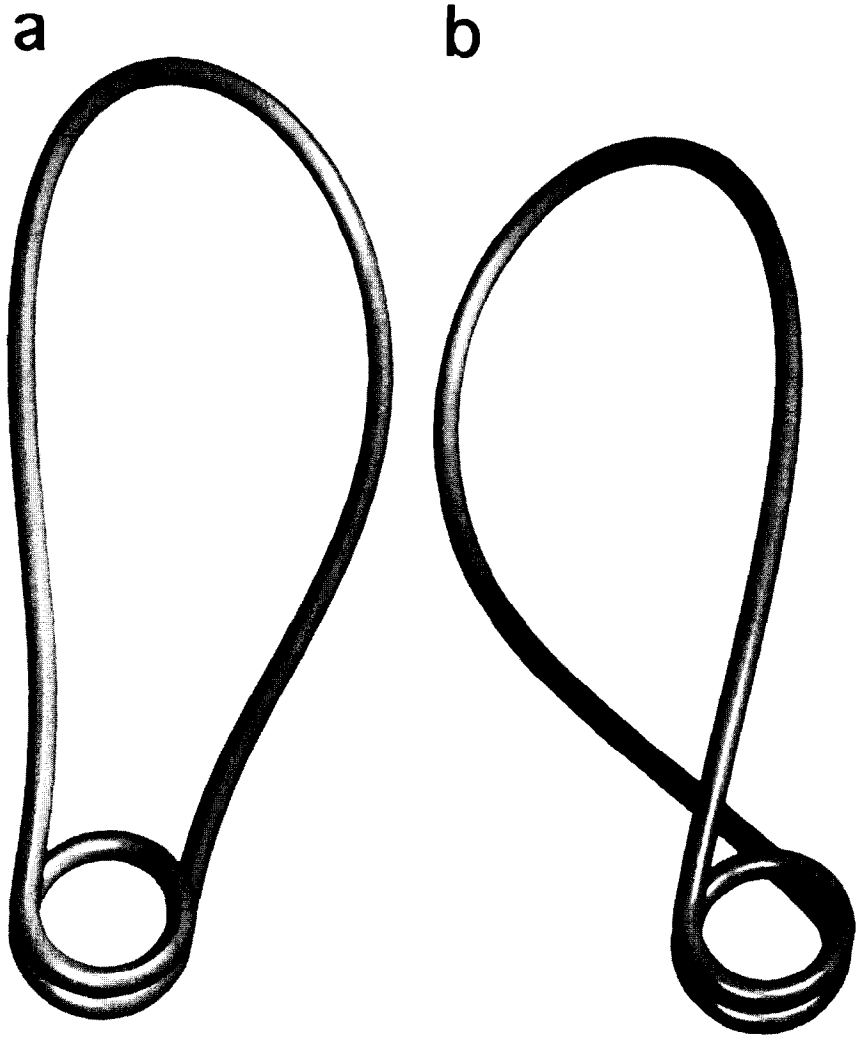

Fig. 8. Molecular representations of the sharp configurational jump associated with increased wrapping of a 500 bp torsionally relaxed DNA around an idealized 'phantom' protein. (a) Open configuration ($Wr = -1.0$) where a segment is wrapped $T = 1\frac{1}{2}$ superhelical turns around the cylindrical protein core; (b) collapsed state ($Wr = -1.6$) where the wrapping is increased (by ca. 10 bp) to $1\frac{5}{8}$ turns. To facilitate comparison, the protein-bound segments of the two structures are drawn in identical orientations. See legend to Fig. 7 for further details.

Fig. 9. Molecular illustration of the effects of chain length on the global equilibrium configuration of a nicked DNA circle with two protein-bound segments, each ca. 140 bp in length and wrapped $1\frac{3}{4}$ superhelical turns around an idealized protein core. Bound segments in the upper-right of each structure are superimposable. (a) 518 bp DNA ($Wr = -2.5$) with extended linker arms of 75 and 163 bp; (b) 1216 bp DNA ($Wr = -2.6$) with one open short (75 bp) and one collapsed long (861 bp) spacer. See text for additional details.

and separated by a contour length of ca. 75 bp. The axes of the free DNA and the bound proteins are represented by piecewise B-spline functions, a localized curve-fitting technique frequently employed in simulations of supercoiled DNA [65,66,72,80,81]. (In the case of B-splines, the helical axis is made up of a series of smooth piecewise curves defined by a set of controlling points, which are optimized in the course of computer simulation.) The bending constant in the examples in Fig. 9 corresponds, as above, to a 5.8° per residue bending fluctuation at 298 K. The configuration of the unbound DNA spacer in the two chains is consistent with theoretical expectations. The shorter spacers are nearly fully extended in both chains, whereas the long spacer in the 1216 bp chain forms a closed loop with a point of self-contact and terminal base pairs oriented at an angle of 154°. The similar writhing numbers of the two DNAs, however, reflect more than this difference in loop folding. There are close contacts between the two spacers in the shorter chain which raise the magnitude of its writhing number (-2.5) close to that (-2.6) of the longer DNA.

10. Protein-induced structural effects

When the torsionally relaxed DNAs in Fig. 8 are ligated and bound protein is removed, the two chains adopt different global equilibrium states. The chain wrapped $1\frac{1}{2}$ superhelical turns (where ΔLk is now -1) relaxes to a circle and that wrapped $1\frac{5}{8}$ turns (where $\Delta Lk = -1.7$) to a figure-8 if the ratio of bending to twisting force constants A/C is taken as unity [73]. The sudden jumps in Wr with T may thus be important to chemical mechanisms by which supercoils are added and removed to and from DNA. These observations provide a first step in understanding how small fluctuations in the local structure of a DNA–topoisomerase complex can be translated into large-scale macromolecular changes. The calculations are consistent with the existence of small chemical agents and/or environmental factors that facilitate or inhibit minute degrees of DNA binding and thus influence the action of these proteins [69].

The degree of protein binding further determines which residues, normally distant along the polymer contour, are brought into direct interaction and also defines the orientation of nonbonded chain segments [69]. This rather subtle long-range control is relevant to the action of other kinds of DNA-binding proteins, allowing communication between spatially separate elements on the chain. Geometric changes at the ends of the free DNA loop control the positioning of

intervening residues with particular restrictions on the protein boundary fixing specific nonbonded sites at characteristic distances and orientations. It is thus easy to understand why the activity of a regulatory protein may be linked to the binding or reactivity of other molecular agents at different sites along the DNA. As illustrated below, the ca. 120° self-crossing angle of free duplex in a torsionally relaxed circular DNA wrapped $1\frac{5}{8}$ turns around a cylindrical protein core positions distant nucleotides on one strand so

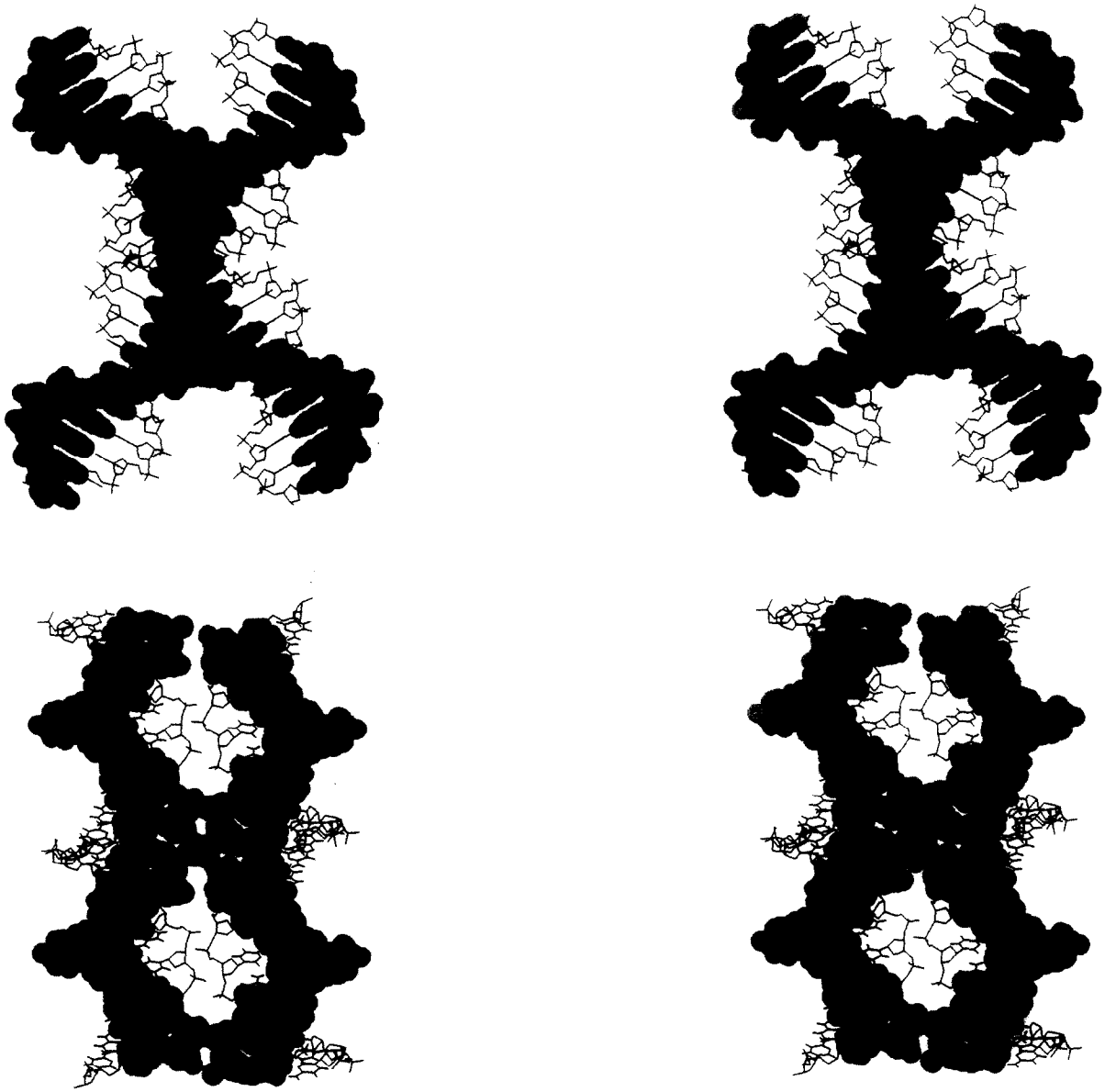

Fig. 10. Color-coded atomic representation of the long-range contacts of free DNA in a circular chain constrained to a 120° orientation by left-handed wrapping (ca. $1\frac{5}{8}$ turns) around a cylindrical protein (superhelical radius $r = 45$ Å, pitch $p = 30$ Å). Contacted strands shown in two shades of orange on the left are chemically exchanged in the images on the right. The resulting cross-over strands are shown in red and yellow. The top views, which correspond to the optimized arrangement of the free DNA loop shown in Fig. 8b, are rotated about the vertical axis by 90° to generate the views shown in the bottom row. The unexchanged circular strand is shown in blue in all cases.

that they can exchange sequential order and form a Holliday junction, a four-armed branched structure implicated in genetic recombination [82]. Both the angle of contact and the geometric feasibility of Holliday structure formation are altered when the DNA is wrapped a different number of turns around the protein core [83]. Thus, the specific geometry of binding one protein to DNA can potentially affect the action of a second protein.

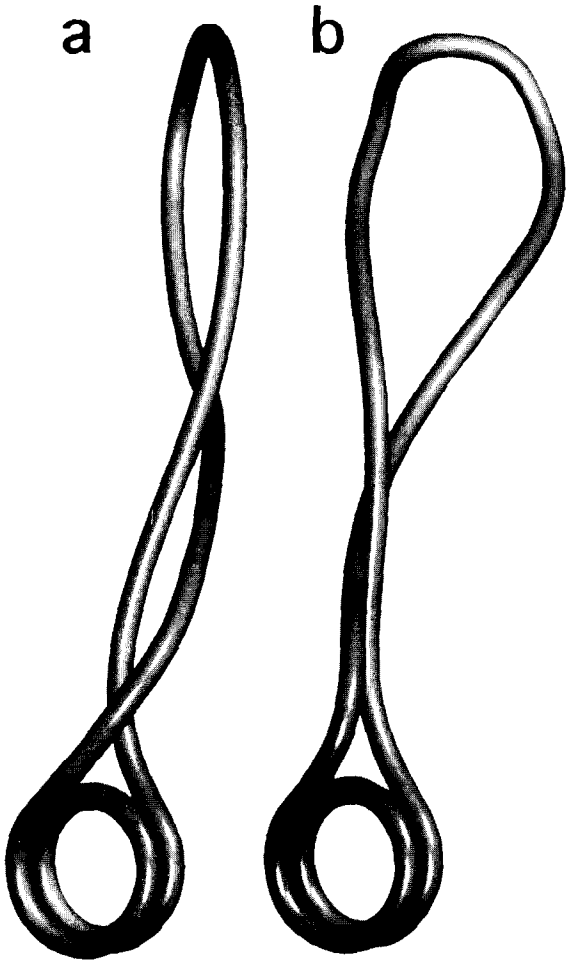

Fig. 11. Illustrative examples of the structural asymmetry of (a) negatively ($\Delta Lk = -4$) versus (b) positively ($\Delta Lk = 1$) supercoiled DNA chains containing a bound protein. Structures are presented so that the ca. 140 bp segments wrapped $1\frac{3}{4}$ turns around the idealized cylindrical core (with superhelical radius $r = 45$ Å and pitch $p = 30$ Å) are superimposable. Total chain length is 500 bp.

Shown in Fig. 10 are atomic representations of the contact zone of the free DNA in a circular chain constrained by wrapping ca. $1\frac{5}{8}$ superhelical turns around a cylindrical protein, prior to and following the exchange of phosphate linkages on distant parts of the same strand. The two structures are identified in modeling studies that combine a backbone generating algorithm with hard sphere packing calculations [83]. Two views are shown, that on the top correspond to the strand crossing arrangement of the free DNA loop seen in Fig. 8b and those on the bottom obtained by rotating the corresponding upper views 90° about the vertical axis. The unexchanged strand is shown in blue in both cases and the exchanged sequences in different colors. Before chemical exchange the contacted strands (shown in two shades of orange on the left of the figure) run antiparallel to the unexchanged strands, while afterward (right side of figure) they make two sharp turns of 120° (one illustrated in red and the other in yellow). Base stacking and overall positioning of DNA are identical in the two structures, although following chemical exchange and relegation, the complex is made up of three independent species — the intact unexchanged circular strand and two smaller single-stranded fragments of exchanged DNA.

The presence of bound protein on closed DNA also influences the uptake of supercoiling. The optimized structures reported in Fig. 11 are supercoiled chains of 500 bp wrapped $1\frac{3}{4}$ left-handed superhelical turns around the protein core described above [69]. As pointed out by Eq. 14, the twist energy of chains with a linking number deficit or excess is deduced in the isotropic rod approximation from the square of the difference between the imposed ΔLk and the writhing number of the helical axis. In these examples, the A/C is taken as 1.35 and ΔLk is -4 or $+1$. The total energies of the two states (3.4 kT and 3.6 kT, respectively) suggest that it may be more difficult to introduce a linking number excess than a linking number deficit in the DNA partially bound to protein. (The magnitude of the change in ΔLk with respect to the twist-free state at $\Delta Lk = -1.7$, however, is not identical in these two examples). The constraints on the ends of the free DNA apparently make it more difficult for the loop to adopt an interwound form with duplex strands intertwined in a

left-handed fashion compared to a configuration with the distant parts of the double helix contacted through right-handed wrapping of duplex strands. This idea is supported by the molecular representation in Fig. 11b which shows local bending at the ends of the protein binding site in the $\Delta Lk = 1$ structure apparently in opposition to the intertwining of the free DNA chain segment. There is no such bending at the junction between free and protein-bound DNA in the $\Delta Lk = -4$ structure in Fig. 11a. The structural differences in these two complexes may be relevant to the competing effects of different proteins on closed circular chains. On purely elastic grounds based on these calculations, it should be easier to add an additional nucleosome to a DNA minichromosome with one bound histone core than to add a protein with a right-handed superhelical binding site, which would increase ΔLk.

11. Equilibrium equations of real DNA

In order to account for base sequence effects in supercoiled DNA, it is necessary to introduce a coordinate transformation that relates the local frame, $[\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3]$, of a base pair to the global coordinate system, $[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$, of the DNA rod. The axes of the local frame are here defined in terms of Euler angles (ϕ, θ, ψ) using the matrix transformation, $\mathbf{d}_i = \mathbf{T}(\phi,\theta,\psi)\,\mathbf{e}_i$, where $i = 1,3$ and $T(\phi,\theta,\psi)$ is given by:

$$\mathbf{T}(\phi,\theta,\psi) = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \cos\phi\cos\psi - \sin\phi\cos\theta\sin\psi & \sin\phi\cos\psi + \cos\phi\cos\theta\sin\psi & \sin\theta\sin\psi \\ -\cos\phi\sin\psi - \sin\phi\cos\theta\cos\psi & -\sin\phi\sin\psi + \cos\phi\cos\theta\cos\psi & \sin\theta\cos\psi \\ \sin\phi\sin\theta & -\cos\phi\sin\theta & \cos\theta \end{bmatrix} \quad (15)$$

The Euler angles and their derivatives with respect to arc length (ϕ', θ', ψ') are also related to the bending components and the twist of the local coordinate frame [67]. If coordinate axes are assigned to every base pair, the roll, tilt, and twist angles can be approximated by:

$$\begin{bmatrix} \rho = (\theta'\cos\psi + \phi'\sin\theta\sin\psi)\Delta s \\ \tau = (-\theta'\sin\psi + \phi'\sin\theta\cos\psi)\Delta s \\ \Omega = (\phi'\cos\theta + \psi')\Delta s \end{bmatrix} \quad (16)$$

Constraints on the vectorial position of the ends of the rod are introduced through a Lagrange multiplier $\mathbf{L} = [\lambda_1, \lambda_2, \lambda_3]$, which acts as the effective force needed to keep the chain ends at the desired end-to-end location, $\mathbf{r}$. The value of $\mathbf{L}$ is initially unknown but can be determined along with the set of ρ, τ, Ω at each base pair for a given chain configuration. The end-to-end vector is taken as the sum of virtual bond vectors $\mathbf{v}$ of length Δs along successive base pair normals and is expressed in terms of the Euler angles as,

$$\mathbf{r} = \sum_{j=1}^{N} \mathbf{v}_j \quad (17)$$

where $\mathbf{v}_j = [\sin\phi_j \sin\theta_j, -\cos\phi_j \sin\theta_j, \cos\theta_j]\,\Delta s$.

The set of Euler angles and the components of $\mathbf{L}$ that minimize the energy of the constrained DNA must satisfy three Euler–Lagrange equations of the form,

$$\frac{\mathrm{d}}{\mathrm{d}s}\frac{\partial\left[\frac{A_1}{2}(\rho-\rho_0)^2 + \frac{A_2}{2}(\tau-\tau_0)^2 + \frac{C}{2}(\Omega-\Omega_0)^2 + \mathbf{L}\cdot\mathbf{v}\right]}{\partial\xi'} = \frac{\partial\left[\frac{A_1}{2}(\rho-\rho_0)^2 + \frac{A_2}{2}(\tau-\tau_0)^2 + \frac{C}{2}(\Omega-\Omega_0)^2 + \mathbf{L}\cdot\mathbf{v}\right]}{\partial\xi}, \quad (18)$$

where $\xi = \phi$, θ, and ψ and ξ' is the derivative with respect to arc length. These expressions lead to a system of $3N$ second-order nonlinear differential equations that can be solved numerically. Details of their solution will be reported elsewhere (T. Westcott, I. Tobias, and W. Olson, manuscript in preparation). The elastic equilibrium structure of a real DNA can also be identified with finite element techniques, as described recently [67,84].

12. Configurational properties of constrained real DNAs

The preferred configurations of closed DNAs with realistic chemical features are strikingly different from those of homogeneous naturally straight rods. As shown below, intrinsically curved DNA fragments not only adopt different three-dimensional arrangements from chains that obey the isotropic rod

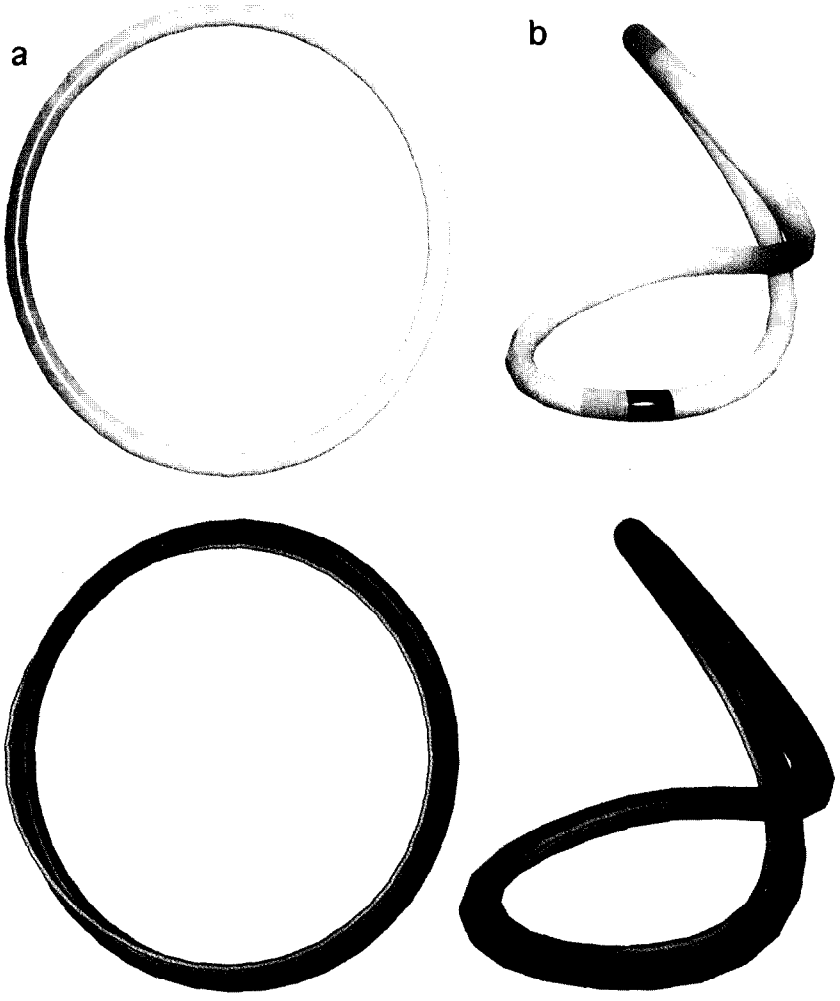

Fig. 12. Color-coded representations of the variation in local twist along the chain contour of: (a) circular and (b) collapsed configurations of a naturally curved 1000 bp DNA with an imposed linking number deficit $\Delta Lk = -1$. Regions of under twisting are colored in the upper structures in shades of red (maximum saturation at $\Delta\Omega = -1°/\text{bp}$), states of over twisting in shades of violet (full saturation at $\Delta\Omega = +1°/\text{bp}$), and residues with zero twist in yellow. The light blue line in the models on the bottom half of the figure corresponds to the end of the curvature vector in the torsionally relaxed state where $\Delta Lk = 0$. See text for additional details.

model but also show pronounced variations in the local twist of adjacent base pairs. Bending anisotropy similarly affects the optimum shape and the distribution of local twist.

If DNA is intrinsically curved (i.e., $\langle \rho \rangle \neq 0$ and/or $\langle \tau \rangle \neq 0$), the excess twist density imposed upon supercoiling depends upon position along the chain [85]. As noted above, twist is uniform in a naturally straight isotropic rod. Illustrated in Fig. 12 are color-coded representations of the variation of twist along the chain contour of two configurations identified in Monte Carlo studies of a 1000 bp naturally curved supercoiled DNA (S. Pedersen, unpublished data). The path of the helical axis is generated at random and the distribution of helical twist is obtained by solving the Euler–Lagrange equations for the given configuration. The linking number deficit is identical ($\Delta Lk = -1$) in the two configurations illustrated here, as are the bending and twisting rest angles (0.36° and 36° per residue, respectively) and the root-mean-square angular fluctuations (5.8° and 4.8° in bend and twist, respectively). In the relaxed twist-free state, the chain would close naturally into a 1000 bp circle made up of 100 complete double helical turns. The collapsed figure-8 structure in Fig. 12 with writhing number -0.9 and total elastic energy 4.6 kT, however, is closer to the global minimum than the circular form with 8.5 kT total energy under the imposed linking number deficit. As evident from the color-coding in the representations on the left-hand side of the figure, the twist is not uniformly distributed in the two structures. Regions of under twisting are colored in shades of red (maximum saturation at $\Delta\Omega = -1°$/bp), states of over twisting in shades of violet (full saturation at $\Delta\Omega = +1°$/bp), and residues with zero twist in yellow. The total twist deformation is greater in the circle ($-360°$ spread out over 1000 bp) than the collapsed state where $Wr \approx \Delta Lk$ and the total twist is near zero. There are, nevertheless, sizable fluctuations from the average twist per base pair in both the circle (with values of $\Delta\Omega$ ranging from -0.83 to $-0.09°$/bp compared to a mean value of $-0.36°$/bp) and the collapsed form (where $\Delta\Omega = -0.29$ to $0.28°$/bp versus $\langle \Delta\Omega \rangle = -0.001°$/bp).

The light blue line in the models on the bottom half of Fig. 12 corresponds to the end of the curvature vector in the torsionally relaxed configuration, or the parts of the DNA that lie on the inside of the naturally formed circle when $\Delta Lk = 0$. Its variation

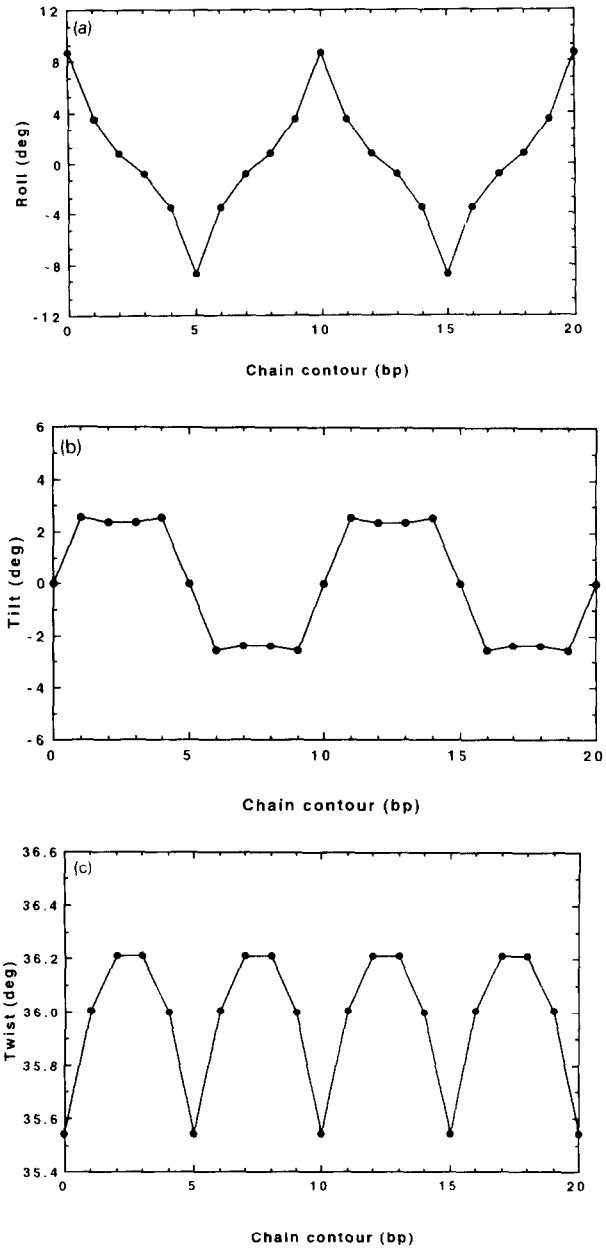

Fig. 13. Variation of: (a) roll ρ, (b) tilt τ, and (c) twist Ω angles with chain contour in a torsionally relaxed 80 bp circular DNA constrained to a plane and subject to anisotropic bending. Force constants impeding rolling, tilting, and twisting (A_1:A_2:C = 1:3.2:1) are taken from Fig. 1. Because the pattern of angles is repetitive, data are reported for only part of the chain.

in the circle upon supercoiling is in agreement with predictions of elasticity theory [85] and also consistent with recent finite element analyses of naturally curved DNA [84]. As is clear from the path of the light blue line, the twisting brought about by decrease in ΔLk brings selected residues from the inside to the outside of the circle. In other words, some residues normally protected on a DNA wrapped in a circle around a protein would become accessible to chemical probes if ΔLk were altered but the duplex pathway preserved. The computed path of the light blue line in the collapsed figure-8 also reproduces the predictions of elasticity theory, a check routinely made of the numerical data. By folding into the collapsed form, residues exposed on the supercoiled circle return to the inside of the structure. As evident from the violet shading in parts of the collapsed structure (upper figure), nearly half of the residues are overwound so that the light blue line lies on the inside of the curve.

A nicked intrinsically straight DNA obeying the isotropic rod approximation bends and twists uniformly into a perfect circle when one of its ends is chemically linked. The local geometry of the chain, however, is no longer uniform if the DNA bends in an anisotropic manner, such as rolling more easily than tilting. Such a DNA bends sharply and preferentially along the roll axis of selected base pairs (T. Westcott, unpublished data). Because the 80 bp contour length of the chain described in Fig. 13 is a multiple of the assumed double helical repeat (10 bp/turn), this DNA is strongly kinked at half helical turns (i.e., every five base pairs). The equilibrium structure, identified by application of the Newton–Raphson method to solve the Euler–Lagrange equations, is closer to a 16-sided polygon than a circle. Here the ratio of force constants, A_2/A_1, impeding tilting versus rolling is taken as 3.2 and that hindering roll compared to twist, A_1/C, is set at 1.0, corresponding to the ratios of observed fluctuations in roll, tilt, and twist in the B-DNA crystal data represented in Fig. 1. As a result of these preferences, the chain tends to relieve the elastic stress associated with ring closure through large changes in roll angle (ca. 9°; Fig. 13a) at points where the base pairs can bend into either the major or minor groove. The tilt angle is more constant, fluctuating in a regular pattern between values of $\pm 2.5°$ (Fig. 13b). These local variations closely mimic the bending tendencies in linear DNA predicted by Zhurkin and co-workers [45] on the basis of the conformational energies of isolated base pairs, with the roll angle alternating between positive and negative values every five base pairs along the DNA. Here in the circular elastic rod model of DNA we find small accompanying decreases (ca. $-0.4°$) in twist at points of large base pair roll and slight increases in twist (ca. 0.2°) at other steps (Fig. 13c).

13. Discussion

Understanding the effects of local base sequence in long DNAs requires computational techniques that link the specific bending, twisting, and translational fluctuations of individual nucleotide residues to average macroscopic properties. Several such approaches have been presented in this review. Applications of polymer chain statistics to linear DNA (a computational methodology for evaluating statistical mechanical average properties of a macromolecule in terms of the constituent atoms and bonds) show how the sequence dependent bending and twisting of neighboring base pairs influence the configuration of open duplexes and how small perturbations of local structure translate into large scale effects at the macromolecular level (Figs. 3–5). Structural data characterizing random sequence B-DNA dimer steps in nucleic acid crystals show surprisingly good agreement with the apparent flexibility of the double helix in dilute aqueous solution [24]. Rules describing the equilibrium rest states and fluctuations of individual base steps are only beginning to be deciphered from the X-ray data.

Alternatively, one can start with a three-dimensional molecular structure and examine effects of local conformational changes on the chain as a whole [65,66,68,73,80,81]. This is a particularly useful way to study spatially constrained DNA polymers, such as closed circles and loops. The size of the problem (tens or hundreds of thousands of atoms in a typical plasmid) together with the computational difficulties of preserving chain closure necessitate simplified representations of the DNA. We have taken advantage of curve fitting methods commonly used in computer-aided design and engineering to generate

spatially constrained configurations and have routinely simplified the representation of chain structure to a single virtual bond level (i.e., one point per residue). Until recently, the long double helix has been treated simply as a symmetric, linearly elastic rod which, like a guitar string, exhibits no preferential directions of bending and twisting. Newly developed numerical techniques now permit the analysis of local sequence-dependent bending and twisting preferences in spatially constrained DNAs.

The presence of a bound ligand has been treated by introducing local geometric parameters in the DNA that mimic the interaction of interest. Protein–DNA complexes have been simulated, for example, by assigning the requisite (fixed) geometry to the protein-bound DNA segment. Small variations in the extent of protein association can translate into large-scale macromolecular effects, transforming a hairpin loop into a collapsed form with a point of long-range self-contact (Fig. 8). These large-scale changes suggest new models for the mechanism of action of enzymes which add or remove supercoils from DNA. Large-scale configurational changes of DNA do not necessarily need to proceed via topological mechanisms whereby DNA is cleaved so that one chain segment can pass through the opening, but rather can take advantage of slight adjustments at the ends of a protein binding site which cause the adjacent free DNA to flip [69,79]. The spacing and geometry of binding sites become important factors in determining overall DNA structure when more than one protein is bound to the double helix. As illustrated in Fig. 9, closely spaced residues alter structure differently from widely separated species. Differences in the torsionally relaxed state translate into distinctly different geometries upon supercoiling (e.g., the preference for branched versus linear supercoil configurations is related to the spacing between binding sites). Competition between protein binding sites on a closed DNA can potentially lead to large-scale changes in three-dimensional structure.

The incorporation of local sequence dependent structural effects introduces further complexity in the treatment of supercoiled DNA. Intrinsic curvature and anisotropic bending at the base pair level are linked to nonuniform changes in local helical twist. Residues that might be less accessible on a naturally straight DNA closed into a circle may become more exposed in a supercoiled chain containing regions of natural curvature. Chain bending that occurs preferentially along a particular base pair axis is also tied to the intrinsic helical twist, with points of strongest bending occurring when the favored axis is aligned with the bending direction. In other words a short circular duplex with a given helical repeat will show sharp bends every half helical turn (e.g., every 5 base pairs in a tenfold double helix; Fig. 13). The bends, however, will be less pronounced in longer chains where the deformation needed to superimpose chain ends is smaller. Since changes in local twist can substantially affect the binding of proteins to DNA, the affinity of a protein for a specific sequence can vary with superhelicity and at different sites along a supercoiled DNA, even if the fragments are identical to one other in the relaxed stress-free state. Thus, one can begin to see how the combination of sequence-directed curvature and supercoiling can facilitate the recognition of specific parts of a DNA.

Acknowledgements

We are grateful to Ms. Fannie Huang for technical assistance with database searches and Mr. Andrew Olson for dedicated help with computer graphics. We are also indebted to our close collaborators who have worked with us in studying DNA — Professor Helen M. Berman (Rutgers University) for assistance with nucleic acid crystal analysis, Professor Tamar Schlick (New York University) for mathematical techniques used to study supercoiled DNA, and Dr. Victor B. Zhurkin (NIH) for insights into DNA conformation and flexibility. This research has been generously supported by the US Public Health Service under research grants GM20861 and GM34809. Calculations were performed at the Rutgers Center for Computational Chemistry, the Pittsburgh Supercomputer Center, and through the facilities of the Nucleic Acid Database project (NSF grant DIR 9012772). J.A.M. is a Howard Hughes Medical Institute predoctoral fellow. S.C.P. and T.P.W. are predoctoral trainees supported in part by grants from the US Public Health Service (Molecular Biophysics Training Grant GM08319) and the US Department of Education (National Needs in Chemistry Program).

References

[1] K.S. Matthews, Microbiol. Rev., 516 (1992) 123.
[2] R. Schleif, Ann. Rev. Biochem., 61 (1992) 199.
[3] R.E. Dickerson, H.R. Drew, B.N. Conner, R.M. Wing, A.V. Fratini and M.L. Kopka, Science, 216 (1982) 475.
[4] R.E. Dickerson, Scientific American, 249 (1983) 94.
[5] O. Kennard and W.N. Hunter, Angew. Chem., Int. Ed. Engl., 30 (1991) 1254.
[6] D. Bhattacharyya and M. Bansal, J. Biomol. Str. Dyn., 10 (1992) 213.
[7] L. Joshua-Tor and J.L. Sussman, Curr. Opinion Str. Biol., 3 (1993) 323.
[8] O. Kennard and S.A. Salisbury, J. Biol. Chem., 268 (1993) 10701.
[9] U. Heinemann, C. Alings and M. Hahn, Biophys. Chem., 50 (1994) 157.
[10] V.B. Zhurkin, N.B. Ulyanov, A.A. Gorin and R.L. Jernigan, Proc. Natl. Acad. Sci. USA, 88 (1991) 7046.
[11] W.K. Olson, N.L. Marky, R.L. Jernigan and V.B. Zhurkin, J. Mol. Biol., 232 (1993) 530.
[12] J.C. Marini, S.D. Levene, D.M. Crothers and P.T. Englund, Proc. Natl. Acad. Sci. USA, 79 (1982) 7664.
[13] H.-M. Wu and D.M. Crothers, Nature (London), 308 (1984) 509.
[14] P.J. Hagerman, Biochemistry, 24 (1985) 7033.
[15] S. Diekmann, FEBS Lett., 195 (1986) 53.
[16] L.E. Ulanovsky, M. Bodner, E.N. Trifonov and M. Choder, Proc. Natl. Acad. Sci. USA, 83 (1986) 862.
[17] J. Griffith, M. Bleyman, C.A. Rauch, P.A. Kitchin and P.T. Englund, Cell, 46 (1986) 717.
[18] G. Muzard, B. Théveny and B. Révet, EMBO J., 9 (1990) 1289.
[19] C.H. Laundon and J.D. Griffith, Cell, 52 (1988) 545.
[20] I. Goulet, Y. Zivanovic, A. Prunell and B. Révet, J. Mol. Biol., 200 (1988) 253.
[21] Y. Zivanovic, I. Goulet, B. Révet, M.L. Bret and A. Prunell, J. Mol. Biol., 200 (1988) 267.
[22] Y. Zivanovic, I. Duband-Goulet, P. Schultz, E. Stofer, P. Oudet and A. Prunell, J. Mol. Biol., 214 (1990) 479.
[23] I. Duband-Goulet, V. Carot, A.V. Ulyanov, S. Douc-Rasy and A. Prunell, J. Mol. Biol., 224 (1992) 981.
[24] A.A. Gorin, V.B. Zhurkin and W.K. Olson, J. Mol. Biol., (1994) submitted.
[25] S.C. Schultz, G.C. Shields and T.A. Steitz, Science, 253 (1991) 1001.
[26] Y. Kim, H.J. Geider, S. Hahn and P.B. Sigler, Nature, 365 (1993) 512.
[27] J.L. Kim, D.B. Nikolov and S.K. Burley, Nature, 365 (1993) 520.
[28] D. Rhodes, Nucleic Acids Res., 6 (1979) 1805.
[29] G.R. Kunkel and H.G. Martinson, Nucleic Acids Res., 9 (1981) 6869.
[30] A. Prunell, EMBO J., 1 (1981) 173.
[31] C.A. Edwards and R.A. Firtel, J. Mol. Biol., 180 (1984) 73.
[32] H.J. Puhl, S.R. Gudibande and M.J. Behe, J. Mol. Biol., 222 (1991) 1149.
[33] J.C. Wang, Proc. Natl. Acad. Sci. USA, 76 (1979) 200.
[34] L.J. Peck and J.C. Wang, Nature (London), 292 (1981) 375.
[35] N.L. Marky and W.K. Olson, Biopolymers, 34 (1994) 109.
[36] N.L. Marky and W.K. Olson, Biopolymers, 34 (1994) 121.
[37] H.M. Berman, W.K. Olson, D.L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A.R. Srinivasan and B. Schneider, Biophys. J., 63 (1992) 751.
[38] M.S. Babcock and W.K. Olson, J. Mol. Biol., 237 (1994) 98.
[39] R.E. Dickerson and H.R. Drew, J. Mol. Biol., 149 (1981) 761.
[40] H.R. Drew, S. Samson and R.E. Dickerson, Proc. Natl. Acad. Sci. USA, 79 (1982) 4040.
[41] A.D. DiGabriele, M.R. Sanderson and T.A. Steitz, Proc. Natl. Acad. Sci. USA, 86 (1989) 1816.
[42] D. Pörschke and M. Eigen, J. Mol. Biol., 5 (1971) 109.
[43] R. Chandrasekaran and S. Arnott in W. Saenger (Editor), Landolt–Börnstein Numerical Data and Functional Relationships in Science and Technology, New Series, Group VII: Biophysics, Vol. 1, Nucleic Acids, Subvolume B, Crystallographic and Structural Data, Springer Verlag, Berlin, 1989 p. 31.
[44] M.S. Babcock, E.P.D. Pednault and W.K. Olson, J. Mol. Biol., 237 (1994) 125.
[45] V.B. Zhurkin, Y.P. Lysov and V. Ivanov, Nucleic Acids Res., 6 (1979) 1081.
[46] E.N. Trifonov and J.L. Sussman, Proc. Natl. Acad. Sci. USA, 77 (1980) 3816.
[47] H.R. Drew and A.A. Travers, J. Mol. Biol., 186 (1985) 773.
[48] S.C. Satchwell, H.R. Drew and A.A. Travers, J. Mol. Biol., 191 (1986) 659.
[49] O. Kratky and G. Porod, Rec. Trav. Chim. Pays-Bas, 68 (1949) 1106.
[50] I.S. Gradshteyn and I.M. Ryzhik, Table of Integrals, Series, and Products, Academic Press, San Diego, CA, 1980.
[51] H. Eisenberg, Acts. Chem. Res., 20 (1987) 276.
[52] P.J. Hagerman, Ann. Rev. Biophys. Biophys. Chem., 17 (1988) 265.
[53] H. Eisenberg, Eur. J. Biochem., 187 (1990) 7.
[54] L. Song, B.S. Fujimoto, P. Wu, J.C. Thomas, J.H. Shibata and J.M. Schurr, J. Mol. Biol., 214 (1990) 307.
[55] D. Shore and R.L. Baldwin, J. Mol. Biol., 170 (1983) 957.
[56] D. Shore and R.L. Baldwin, J. Mol. Biol., 170 (1983) 983.
[57] A.R. Srinivasan, R. Torres, W. Clark and W.K. Olson, J. Biomol. Str. Dyn., 5 (1987) 459.
[58] W.K. Olson, A.R. Srinivasan, R.C. Maroun, R. Torres and W. Clark in R.D. Wells and S.C. Harvey (Editors), Unusual DNA Structures, Springer Verlag, New York, 1987, p. 207.
[59] D. Rhodes and A. Klug, Nature (London), 292 (1981) 378.
[60] R.C. Maroun and W.K. Olson, Biopolymers, 27 (1988) 561.
[61] R.C. Maroun and W.K. Olson, Biopolymers, 27 (1988) 585.
[62] A.R. Srinivasan and W.K. Olson, J. Mol. Graphics, 6 (1988) 126.
[63] W.K. Olson, A.R. Srinivasan and M.-H. Hao in L.P. Kartashev and S.I. Kartashev (Editors), Proc. Supercomputing '89: Supercomputer Applications, International Supercomputing Institute, Inc., St. Petersburg, Florida, 1989, p. 361.
[64] J.H. White, Am. J. Math., 91 (1969) 693.

[65] M.-H. Hao and W.K. Olson, Macromolecules, 22 (1989) 3292.
[66] T. Schlick and W.K. Olson, J. Mol. Biol., 223 (1992) 1089.
[67] Y. Yang, I. Tobias and W.K. Olson, J. Chem. Phys., 98 (1993) 1673.
[68] P. Zhang, W.K. Olson and I. Tobias, Comp. Polymer Sci., 1 (1991) 3.
[69] P. Zhang, I. Tobias and W.K. Olson, J. Mol. Biol., 242 (1994) 271.
[70] M.E. Mortenson, Geometric Modeling, John Wiley, New York (1985).
[71] F.B. Fuller, Proc. Natl. Acad. Sci. USA, 68 (1971) 815.
[72] W.K. Olson and P. Zhang, Methods in Enzymology, 203 (1991) 403.
[73] G. Liu, W.K. Olson and T. Schlick, Comp. Polymer Sci., (1994) submitted.
[74] J.M. Sperrazza, J.C.I. Register and J. Griffith, Gene, 31 (1984) 17.
[75] L. Bracco, D. Kotlarz, A. Kolb, S. Diekmann and H. Buc, EMBO J., 8 (1989) 4289.
[76] C.F. McAllister and E.A. Achberger, J. Biol. Chem., 264 (1989) 10451.
[77] T.J. Richmond, J.T. Finch, B. Rushton, D. Rhodes and A. Klug, Nature (London), 311 (1984) 532.
[78] A. Klug, J.T. Finch and T.J. Richmond, Science, 229 (1985) 1109.
[79] I. Tobias, B. Coleman and W.K. Olson, J. Chem. Phys., 101 (1994) 10990.
[80] T. Schlick and W.K. Olson, Science, 257 (1992) 1110.
[81] T. Schlick, W.K. Olson, T. Westcott and J.P. Greenberg, Biopolymers, 34 (1994) 565.
[82] R. Holliday, Genet. Res. Camb., 5 (1964) 282.
[83] A.R. Srinivasan and W.K. Olson, Biochemistry, 33 (1994) 9389.
[84] W.R. Bauer, R.A. Lund and J.H. White, Proc. Natl. Acad. Sci. USA, 90 (1993) 833.
[85] I. Tobias and W.K. Olson, Biopolymers, 33 (1993) 639.