ELSEVIER

Biophysical Chemistry 55 (1995) 95–114

Biophysical Chemistry

Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature

John A. Schellman [a,*], Stephen C. Harvey [b]

[a] *Institute for Molecular Biology, University of Oregon, Eugene, OR 97403, USA*

[b] *Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, AL 95294, USA*

Abstract

Long molecules of DNA have the statistical properties of a worm-like coil. Deviations from linearity occur both because of small dynamic bends induced by thermal motion and from a random distribution of static bends. The latter originate in the different conformations of each of the possible base pair sequences. In this paper a statistical theory of the persistence length of DNA is developed which includes both static and dynamic effects for each base pair sequence, as well as the sequence-dependent correlations of bending angles. The result applies to a generic DNA, i.e., the average over an ensemble of all possible sequences. The theory is also applied to the generation of the average properties of curved DNAs by an analytic method that includes dynamic averaging as well as correlated bends. These results provide information which supplements that obtained by others using Monte Carlo methods. The additivity relation $1/P = 1/P_s + 1/P_d$ proposed by Trifonov et al., where P is the persistence length and P_s and P_d are the persistence lengths arising from purely static and dynamic effects, respectively, has been verified to be accurate to better than 0.5%. This is true for both a simplified model and one that includes a complete set of static bends at all base pair sequences.

Keywords: Persistence; DNA curvature; DNA bending; Wedge; DNA flexibility; Kinoplast

1. Introduction

It is now generally accepted that DNAs form helices which contain small, systematic, sequence-dependent deformations. These deformations can usually be interpreted as small local bends in the propagating helix axis. The evidence for such bends has accumulated from the study of curved DNAs [1,2] and from crystallography [3]. For many years Trifonov called attention to the possible role such deformations could play when proper phasing permits them to accumulate to an extended curvature [4].

One of the proving grounds for the existence of local bends in DNA is the study of natural and synthetic curved DNAs, which can be observed experimentally by gel separation techniques, hydrodynamics in solution and by electron microscopy. Three theories are current for the origin of the local structural changes which accumulate to observable curvature. The junction model of Wu and Crothers [5] places the bends at the ends of tracts of A or T sequences ($dA_n \cdot dT_n$, $n \geq 4$) which are assumed to form a different type of helix. The wedge model of Trifonov [6] assigns a specific structural bend of 8.7° at each occurrence of the $dA_2 \cdot dT_2$ dinucleotide. Theoretical studies [7–9] and the crystal structure of DNAs containing A-tracts [10,11] indicate that A-

* Corresponding author.

SSDI 0301-4622(94)00144-8

tracts are not curved, leading to a curved general sequence model [7–9,12,13]. This third model is exactly the opposite of Trifonov's, since it proposes that the non-A tracts are the source of curvature, while the A tracts provide proper helical phasing. This issue is as yet unresolved. The present paper makes use of the 'nearest neighbor approximation' [14], deals mainly with long, random sequences of DNA, and will not contribute directly to the problem of determining the correct model.

Another effect of statistical, local bending of DNA helices is the reduction of the persistence length, and other measures of polymer extension in solution. Traditionally, the limited persistence of helical orientation in DNA is assumed to arise from the thermal excitation of helical bending modes. This agrees with the common, but not quite accurate, perception that the persistence length is a measure of chain stiffness. It is in fact a measure of length propagation along a starting axis: the direction of the first bond. It is a memory function, and depends on how far one can go down a random polymer chain before memory of the orientation of the first link is lost. Any kind of bend, stiff or flexible, will result in reduced persistence. As a result permanent structural kinks can reduce the persistence length as well as thermal flexing. This viewpoint has been examined in a recent paper of Trifonov et al. who investigated the effect of small, fixed structural bends (wedges) on the persistence length of DNA and proposed a structural as well as stochastic origin for the persistence length of DNA [15].

In their paper Trifonov et al. performed a Monte Carlo calculation of the effect of fixed bends in the absence of normal flexibility. They assumed a given length of DNA, generated a large number of randomly selected steps of this length, placed an 8.7° bend at each AA and TT sequence, and then calculated the average end-to-end distance of the sample. The calculation was repeated as a function of length and base composition. The chains were straight except for the randomly located fixed bends, so that the only mechanism for chain reorientation was the wedge model discussed above. From the average statistics of these relatively short chains, estimates of the persistence length of long chains were obtained using a formula for the dependence of the persistence length on chain length [16]. This is the equivalent of calculating the average persistence length of a very long DNA molecule with a random sequence, as required in Eq. (2) below. It was found that the persistence length of a random-sequence DNA with an AT/GC ratio of unity was about 635 base pairs. This is about four times the experimental value for the persistence length (which depends on ionic strength). Their conclusion was that a significant, but not dominant, part of the bending of DNA arises from the presence of fixed bends. Finally, using heuristic arguments they proposed that the persistence length, P, of DNA can be decomposed into two parts

$$1/P = 1/P_d + 1/P_s \tag{1}$$

where the terms on the right refer respectively to the persistence based on thermal bending and on structural bending. Trifonov et al. refer to these as 'dynamic' and 'static' persistence lengths. We will also make use of this terminology but we will also emphasize that these two designations are not mutually exclusive. The real difference between a straight linkage and one associated with a wedge is that in the first the equilibrium propagation angle is assumed to be 0° and in the second it is not. Both types of link are flexible in principle.

It should be recognized that one cannot physically define a static persistence length for a fixed sequence of DNA. The static conformation of a sequenced DNA with fixed angles, straight or otherwise, is a three-dimensional structure. There is no limiting persistence length. It is only by averaging over sequences that the static persistence length has meaning. However, it is shown in Appendix A that the persistence length of a heteropolymer is in fact an average over the all sequences that are generated by starting at each unit of the chain and going in both a forward and backward direction. For a long DNA with an unbiassed random sequence this average does converge to give a static persistence length as shown by the calculations of Trifonov et al. and those which are presented below.

As in previous discussions [16–18] we will be representing DNA as a chain of simple structureless vectors which keep track of the local direction of the axis of the duplex structure. This is the virtual bond method [19]. Recently other investigators have started

with dynamic models of the complete atomic structure of DNA [20–24].

The relation of this work to the present approach will come up later in the discussion of results.

2. Analytical calculation of the persistence length

This paper will present an analytical solution for the persistence length of DNA which includes both thermal flexing and kinks or wedges. In the calculations it will be assumed that dynamical bending events are independent of one another, i.e., that the state of bending at a given link does not affect the average angle of bending at other links. In addition the formulation will average over all possible sequences of DNA with a given base composition. Thus the results are for a generic, not a sequenced, DNA. This point of view differs from other treatments of this problem [25] and will be discussed further below.

2.1. Persistence length for heterogeneous chains

For an ordinary homogeneous chain molecule it follows from the theory of worm-like coils that the mean square end-to-end distance, h^2, is related to the persistence length by the formula

$$h^2 = 2PL \tag{2}$$

where P is the persistence length and L is the contour length of the chain. In the appendix it is shown that this formula applies for sequenced, heterogeneous polymers as well, provided P is interpreted as the average of the persistence length using all units as starting points and in both directions. It will be useful to demonstrate by examples the way in which the persistence length depends on direction and position in a heterogeneous system where there is more than one mode of chain propagation. From its definition (Eq. A2) P is given by

$$P = \ell(1 + \langle\cos\theta_{12}\rangle + \langle\cos\theta_{13}\rangle + \langle\cos\theta_{14}\rangle + \cdots)$$

and is evaluated by successive projections down the chain (Fig. 1). ℓ is the length of a link, the height per base pair in our calculation, and $\langle\cos\theta_{jk}\rangle$ is the average of the cosine of the angle between the jth and kth links. If there is no correlation between successive bends and the bending is isotropic, then the evaluation of a typical term consists of a series of nearest neighbor projections. If there is only uniform thermal bending, the only parameter is $\alpha = \langle\cos\theta_{12}\rangle$ and it is easily seen [26] that $\langle\cos\theta_{jk}\rangle = \alpha^{|k-j|}$ (Fig. 1). This converts the above equation to an ordinary geometric series and leads to the standard formula:

$$P = \ell(1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \cdots) = \frac{\ell}{1-\alpha}. \tag{3}$$

On the other hand if there are static or structural contributions to the bending of the chain, which depend on the identity of neighboring base pairs (wedges), the sum will depend on the sequence, and therefore on the location of the starting point and on direction. In papers which start with the atomic structure of DNA and interactomic potential functions, a more complete description of local deformations is obtained in terms of the bending, twisiting and displacement of base pairs. The relationship between the structural work and the virtual bond method has been discussed recently [21,23].

In heterogeneous polymers the persistence length depends on the starting position, the direction, and the contiguous sequence. This is most easily seen with an extreme example. We take a chain with linkages that are either straight, $\langle\cos\theta\rangle = 1$, or at right angles $\langle\cos\theta\rangle = 0$. In this case encountering a

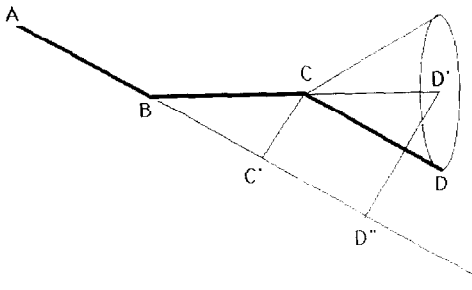


Fig. 1. Isotropic bending. The heavy lines, AB, BC, CD represent three links in the chain. Because the bending is isotropic, the distribution of the link CD possesses cylindrical symmetry relative to BC, indicated by the conical set of loci. CD′ is the mean projection of CD on the axis defined by BC. Since BC is also disposed about AB with cylindrical symmetry, CD′ is subsequently projected on the axis of AB to give C′D″. For unit length this projection is the product of the averaged direction cosines $\langle\cos\theta_1\rangle\langle\cos\theta_2\rangle$.

right-angle link anywhere in a sequence will make the product of projections zero if it includes this link. We take an example of a polymer with a right angle between links 10 and 11 between 20 and 21, and α links everywhere else. Recall that the links are vectors connecting base pairs, not the base pairs themselves. Links are numbered by the number of the lower of the base pairs they connect, regardless of the direction of propagation. Thus P_1^+ refers to the link between base pairs 1 and 2; P_2 starts at the link between base pairs 3 and 2, etc. Then $P_1^+ = 10\ell$; $P_2^- = 2\ell$; $P_5^+ = 6\ell$; $P_{13}^+ = 8\ell$; $P_{13}^- = 3\ell$; etc., where $\pm$ indicates direction in which the persistence is propagated and the subscript labels the starting unit. The unit k is the starting link for both P_k^+ and P_k^-. We are using the notation of the appendix. Note that the series can terminate either by encountering a right-angle unit or by arriving at the end of the chain. It is the average of these possibilities over all the units of the chain as starting points that must be inserted in Eq. (2). For real DNA $\langle \cos\theta \rangle$ will always be a number slightly smaller than unity and the persistence length will not show so marked a dependence on sequential position and direction as in the example.

Initially we consider the earlier Trifonov model with static bends at AA and TT sequences [15]. Later this will be generalized to a full set of static deformations [27].

2.2. Propagation properties of the Trifonov model

Any calculation of the persistence properties of DNA is dependent on the average geometrical properties of adjacent links. Fig. 2, reprinted from Ref. [18], may be used as the representation of an instantaneous bend in DNA. The segment ℓ_β represents the site of the wedge, which is crudely represented as a gap in a broken cylinder. The length ℓ_β can be introduced into the calculation [18], if the actual structure of the bend is known but will be ignored for the present purposes since it will be very small for a small bend between base pairs. Papers which start with structural models explicitly account for this gap, which can include a local translation as well as a bend [23]. The instantaneous angle between units b and m in the figure will be symbolized in general by θ. If θ is normally zero (a dynamic bend) it will be written as θ_d. Bends in DNA are often classified as static (i.e., a fixed wedge angle) or dynamic (no wedge but with an average bend produced by thermal agitation). Since DNA is a conformationally flexible structure, all base pair sequences should be considered as dynamic, so θ will be written as the sum of a static and dynamic component,

$$\theta = \theta_s + \theta_d. \tag{4}$$

θ_s refers to the fixed equilibrium angle of the wedge and θ_d refers to the dynamic or fluctuating thermal displacement. The conformational average $\langle \cos\theta \rangle$ will appear frequently in the analysis. It is given by

$$\langle \cos\theta \rangle = \cos\theta_s \langle \cos\theta_d \rangle + \sin\theta_s \langle \sin\theta_d \cos\phi_d \rangle$$

where the latter term averages to zero if the thermal bend is isotropic about the direction of the fixed bend, or less stringently if the potential energy of the

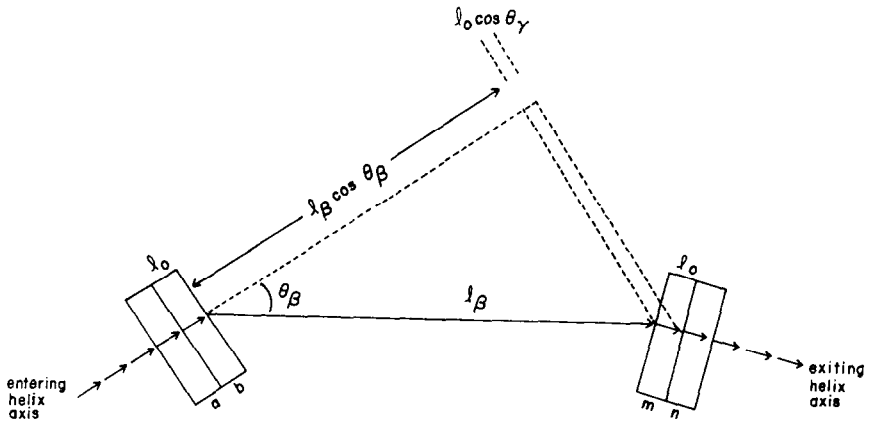


Fig. 2. Representation of a structural bend. See text and Ref. [18] for details.

thermal bend has C2 symmetry. We then have:

$$\langle \cos\ \theta \rangle_s = \cos\ \theta_s \langle \cos\ \theta_d \rangle \tag{5}$$

or

$$\alpha = \alpha_s\, \alpha_d .$$

The second line is a definition of α_s and α_d as the cosine of the fixed wedge angle and the average cosine of the dynamic component, respectively. Actual potentials are calculated in the works of De Santis, Zhurkin, Olson and Beveridge [28–30,21,25,23] so that these averages may be calculated. With the same assumptions it may be shown that the average of the sine of the wedge angle is given by

$$\langle \sin\ \theta \rangle = \sin\ \theta_s \langle \cos\ \theta_d \rangle \equiv \sigma_s\, \alpha_d . \tag{6}$$

For simple models which do not take into account the angular correlations of bends imposed by the helix geometry, α_s and α_d suffice for the calculation of the persistence length.

For isotropic bending with a quadratic dependence of energy on θ_d, $\langle \cos\ \theta_d \rangle$ may be Boltzmann averaged with the result

$$\langle \cos\ \theta_d \rangle = \exp\left[-\Delta\theta_d^2/2 \right] \tag{7}$$

where $\Delta\theta_d^2$ is the mean square thermal fluctuation in θ_d which is related to the bending force constant k_θ by $\Delta\theta_d^2 = kT/k_\theta$ [16]. This presents a better picture of α_d which should be viewed as the Brownian decay of a vector component rather than as the cosine of an angle.

2.3. The AA wedge model

Factors which influence the persistence length when static bends are present are (1) the base composition of the DNA which affects the probability of various pair sequences, (2) the correlation of static bends with the occurrence of specific sequences, and (3) the correlation of the relative orientation of bends determined by their fixed orientation in the base pair framework and the helical geometry which correlates orientations over large sequences. (Static links are not isotropic. If the first AA sequence on the left of AAACGAAT is inclined in the x direction, then the second AA will be at a positive angle of approximately 36°, and the third AA will be at 180°. These correlations, ignored in the first two models, are important in determining the shape of the DNA molecule even when averaged over all sequences).

We will develop our conclusions by adding these factors in steps, one at a time. This will help to make clear the manner in which the above factors enter the theory and test the adequacy of various approximations which have been made in the literature. Finally, the simple AA wedge model will be dropped and a formulation will be presented in which thermally agitated static bends in fixed orientations are associated with all 16 pair sequences.

A primitive result

If one makes the unrealistic assumptions that there is no correlation in link sequence and no correlation in the orientation of wedge angles (bends are isotropic), then the solution to the problem has already been presented as Eq. (5) of Ref. [18]:

$$P/\ell = \frac{1}{1 - p_d \langle \cos\ \theta \rangle_d - p_s \langle \cos\ \theta \rangle_s} = \frac{1}{1 - p_d\, \alpha_d - p_s\, \alpha_d\, \alpha_s} \tag{8}$$

or in more general form

$$P/\ell = \frac{1}{1 - \langle\langle \cos\ \theta \rangle\rangle} \tag{8a}$$

where the double angular brackets indicate averaging over sequence as well as over bending angle. Here p_s is the probability of an AA or TT link between base pairs and is 1/8 for a DNA with a composition of 50% adenine/thymine. p_d, equal to $1 - p_s$, is the probability of an ordinary link. The gap length, ℓ_β in Fig. 2, which should be very small for a narrow wedge, has been taken to be zero. The second form of Eq. (8) shows that the standard formula, Eq. (3), is applicable but that $\langle \cos\ \theta \rangle$ is averaged over both ordinary and bent links. It has been assumed that the stochastic part of cos θ is the same for both static and dynamic bends. We thus assume a symmetric, conical distribution of link vectors around the equilibrium orientation that is the same for bent and straight links. This is not necessary for the theory, but at the present time it is necessary for application to experiments. The formalism can be easily modified, not only to take into account the difference

between bent and straight links, but also a different α_d can be assigned to each type of link if the requisite information is available.

Inverting Eq. (8) and eliminating p_d,

$$\frac{l}{P} = (1 - \alpha_d) + \alpha_d(1 - \alpha_s)p_s. \tag{9}$$

This relation obeys Eq. (1) rather accurately. If there are no wedges, $\alpha_s = 1$ and from Eq. (8), $\ell/P_d = 1 - \alpha_d$, where P_d is the dynamic persistence. This is the first term of Eq. (9). If there are wedges, but no flexibility, then $\alpha_d = 1$ and from Eq. (8) $\ell/P_s = (1 - \alpha_s)p_s$ where P_s is the static persistence. Since α_d differs from unity only by a few thousandths, this is essentially equal to the second term of Eq. (9).

Sequence correlation, but no orientational correlation

The approximations that lead to Eq. (8) are not very good, though it will be seen that they provide a better estimate than one would suspect. The mathematics will now be expanded to evaluate properly the probabilities of sequences of dynamic and static links while the orientation of bends will still be assumed to be uncorrelated. Comparing later results with Eq. (8) or (9) will provide a measure of the importance of the two types of correlation.

The initial problem is to calculate quantities like $\langle \cos \theta_{j,j+n} \rangle$, keeping track of sequence correlations. We do this as before by starting out at link $j + n$ and projecting back one base pair at a time until we arrive at the jth link. The addition of a base pair increases by unity the number of links as well as the number of base pairs. If the bends at each link are considered to be isotropic, the only projections which survive are along the links, i.e., α_d, and $\langle \cos \theta \rangle_s = \alpha_s \alpha_d$ (Fig. 1). The projection of link $j + k$ onto link j is done by forming a product of matrices made up of these projection cosines [19]. The matrix

$$\begin{array}{c} \\ \mathrm{A} \\ \mathrm{T} \\ X \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{A} & \mathrm{T} & X \end{array} \\ \begin{pmatrix} \alpha_d\alpha_s & \alpha_d & \alpha_d \\ \alpha_d & \alpha_d\alpha_s & \alpha_d \\ \alpha_d & \alpha_d & \alpha_d \end{pmatrix} \end{array}$$

$$= \alpha_d \begin{pmatrix} \alpha_s & 1 & 1 \\ 1 & \alpha_s & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

has the property of assigning the correct projection for each type of pair sequence. Since the choice between cytosine and guanine is irrelevant for the present problem, the entries for these have been combined to a single entry X = C or G. The direction cosines must be multiplied by the probability that the added base is A, T, G or C. We will define the probability of A or T by $p_A = p_T$; the probability of X is then $p_X = 1 - 2p_A$. If p_A is 1/4, $p_X = 1/2$. The matrix containing both probabilities and projections is [19]:

$$\mathbf{M} = \begin{pmatrix} p_A\alpha_d\alpha_s & p_T\alpha_d & p_X\alpha_d \\ p_A\alpha_d & p_T\alpha_d\alpha_s & p_X\alpha_d \\ p_A\alpha_d & p_T\alpha_d & p_X\alpha_d \end{pmatrix} \tag{10}$$

Inspection will show that this matrix allots a probability of $+p_A^2$ to pair sequences consisting only of A and T, $p_A p_X$ to sequences containing X and A or T and p_X^2 to XX sequences.

The projection down n links is

$$\langle \cos \theta_{j,j+n} \rangle = (p_A, p_T, p_X)\mathbf{M}^n \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \tag{11}$$

The left row vector takes care of the initial base in the sequence and the right column vector collects all the terms which represent the projections and the probabilities for each possible sequence. The persistence length is the sum of all such projections (n = 1, ∞) plus the first link. This can be written as

$$P/\ell = (p_A, p_T, p_X)\left(\mathbf{I} + \sum_{k=1}^{\infty} \mathbf{M}^k\right)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= (p_A, p_T, p_X)\left(\frac{1}{\mathbf{I} - \mathbf{M}}\right)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{12}$$

where $\mathbf{I}$ is the unit matrix. Note that, just as in Eqs. (3) and (8), we are summing a geometric series, though the element of the series is now a matrix.

The matrix $\mathbf{I} - \mathbf{M}$ is 3×3 and can be inverted analytically for an algebraic solution to the problem. Assuming $p_A = 1/4$ the solution may be written

$$\ell/P = 1 - \alpha_d + \frac{\alpha_d(1 - \alpha_s)}{8 + \alpha_d(1 - \alpha_s)} \tag{13}$$

after some tedious rearrangements. Eqs. (9) and (13) are essentially identical for practical cases. $p_A = 1/4$ in Eq. (10) is equivalent to $p_s = 1/8$ in Eq. (9). According to current data α_s and α_d are both of the order of 0.99. With this value the second term of the denominator in (13) can be neglected with negligible error, leading to a complete equivalence of the two results. Evidently, sequence correlations without geometric correlations have a negligible effect on the static persistence length of DNA.

The AA wedge model with sequence and angular correlation

We now consider the complete problem using geometric-stochastic matrices of the kind discussed in Flory's book [19]. In general one needs a matrix which will project a vector in the coordinate system of a link onto the coordinate system of the preceding link. The total transformation of a vector in unit $j+n$ into the coordinate system of j is then a product of n such matrices.

For static bends the projections on the x and y-axes of a neighboring base pair unit no longer cancel out on averaging as in Fig. 1. Dynamic or static bending will then project these components onto the z component of the next base pair, etc. In addition, if the mathematics is a faithful reflection of the physical system, properly phased wedges will introduce real curvature in the DNA and this curvature will be represented by non-vanishing x and y components.

A local coordinate system must be chosen and we use that of Bolshoy et al. [27], which has also been used by one of us in an earlier work [31]. The z-axis will be chosen as the local helix axis as estimated for straight DNA. Duplex DNA has two sets of pseudo-dyad axes [31]. The standard set lies within the base pair and interchanges the nucleotides of the base pair. The second set lies between base pairs and is generated from the first set by half the helical screw operation. These screw axes interchange neighboring base pairs. We will define the local x-axis to be along the second type of dyad axis with positive x on the side of the major groove [32].

The orientational angles for bending can now be read directly from the table of Bolshoy et al. [27]. In place of each of the scalar elements in the matrix of Eq. (10), we now require a 3 by 3 matrix to transform vectors in the three dimensional space of the link vectors. This complete matrix takes the form

$$\langle \mathbf{M}_g \rangle = \begin{array}{c} \\ A \\ T \\ X \end{array} \begin{array}{c} \begin{array}{ccc} A & T & X \end{array} \\ \begin{pmatrix} p_A\langle \mathbf{M}_{AA} \rangle & p_T\langle \mathbf{M}_d \rangle & p_X\langle \mathbf{M}_d \rangle \\ p_A\langle \mathbf{M}_d \rangle & p_T\langle \mathbf{M}_{TT} \rangle & p_X\langle \mathbf{M}_d \rangle \\ p_A\langle \mathbf{M}_d \rangle & p_T\langle \mathbf{M}_d \rangle & p_X\langle \mathbf{M}_d \rangle \end{pmatrix} \end{array}$$

$$p_X = 1 - 2p_A \qquad (14)$$

This is a 9×9 matrix. The 3×3 blocks pertain to the sequence of A, T and X as before. Individual rows and columns refer to Cartesian coordinates. For example the element (2, 6) refers to the y component of A and the z component of T in an AT sequence. The p factors introduce the probability of an A, T or X at any location in the sequence. The three types of submatrices are for the static and dynamic bends at AA and TT and for the dynamic bends at all other sequences. Each pair sequence of bases is associated with the correct matrix for its geometric propagation.

The matrix we require for a dynamic or thermal bend, $\mathbf{M}_d$, has some subtle properties. It must allow for bends to occur stochastically at an angle ϕ, which is a destruction of orientational memory of the links, yet keep strict track of the helical geometry so that the wedge bends will occur properly phased with respect to one another.As shown in Appendix B the matrix for random bending at a link is given by

$$\langle \mathbf{M}(\theta_d) \rangle = \begin{pmatrix} (1+\alpha_d)/2 & 0 & 0 \\ 0 & (1+\alpha_d)/2 & 0 \\ 0 & 0 & \alpha_d \end{pmatrix}. \qquad (B.3)$$

To obtain $\langle \mathbf{M}_d \rangle$ we multiply (B.3) by the matrix for a helical rotation of angle ϕ_h about the local z-axis,

$$\langle \mathbf{M}_d \rangle = \begin{pmatrix} C(1+\alpha_d)/2 & -S(1+\alpha_d)/2 & 0 \\ S(1+\alpha_d)/2 & C(1+\alpha_d)/2 & 0 \\ 0 & 0 & \alpha_d \end{pmatrix} \qquad (B.5)$$

where $C = \cos\phi_h$ and $S = \sin\phi_h$. It should be noted there is accumulated evidence for a dependence of ϕ_h on pair sequence and values have been proposed for all ten distinct types of pair-sequences. This variability is ignored in the present calculation, but is taken into account in the next two sections.

The matrices $\langle \mathbf{M}_{AA} \rangle$ and $\langle \mathbf{M}_{TT} \rangle$ contain the following operations: (1) generating the average dynamic bending component, (2) locating the angle, ϕ_s, where the bend is to take place, (3) bending by the angle θ_s, and (4) rotating by the helix angle. These steps and the resulting complete matrix, (B.9), are outlined in Appendix B.

$$\mathbf{M}_{ij} = \mathbf{M}_h \cdot \begin{pmatrix} \frac{1+\alpha_d}{2}\left(c\theta_s \cdot c\varphi_s^2 + s\varphi_s^2\right) & \frac{1+\alpha_d}{2} c\varphi_s \cdot s\varphi_s (c\theta_s - 1) & \alpha_d \cdot c\varphi_s \cdot s\theta_s \\ \frac{1+\alpha_d}{2} c\varphi_s \cdot s\varphi_s (c\theta_s - 1) & \frac{1+\alpha_d}{2}\left(c\varphi_s^2 + s\varphi_s^2 \cdot c\theta_s\right) & \alpha_d \cdot s\varphi_s \cdot s\theta_s \\ -\frac{1+\alpha_d}{2} c\varphi_s \cdot s\theta_s & -\frac{1+\alpha_d}{2} s\varphi_s \cdot s\theta_s & \alpha_d \cdot c\theta_s \end{pmatrix} \tag{B.9}$$

where $\mathbf{M}_h$ is the matrix for the helix operation.

The matrix, $\langle \mathbf{M}_9 \rangle$, can now be used to generate the persistence length exactly as in the previous section except for appropriate changes in the end vectors:

$$P/\ell = \mathbf{lv} \cdot \left(\mathbf{I}_9 + \sum_{k=1}^{\infty} \langle \mathbf{M}_9 \rangle^k \right) \cdot \mathbf{rv}$$

$$= \mathbf{lv} \cdot \left(\frac{1}{\mathbf{I}_9 - \langle \mathbf{M}_9 \rangle} \right) \cdot \mathbf{rv} \tag{15}$$

where $\mathbf{I}_9$ is the 9×9 identity matrix, **lv** is $(0, 0, p_A, 0, 0, p_T, 0, 0, p_X)$ and **rv** is $(0, 0, 1, 0, 0, 1, 0, 0, 1)^t$ where t indicates transpose. The forms of **lv** and **rv** stem from the fact that for the persistence problem we require the projection of the z component of one base pair unit onto the z component of another. Though Eq. (15) is the final solution for the AA wedge model, it is not practicable to present it in analytical form. An algebraic inversion of $(\mathbf{I}_9 - \langle \mathbf{M}_9 \rangle)$ produces expressions containing more than 300 000 terms. One can evaluate Eq. (15) for each set of values of the parameters, or one can evaluate all but one of the parameters and invert the matrix with only one variable, since this produces a ratio of polynomials in the selected parameter of order no higher than the ninth.

Calculated results and discussion of the AA wedge model

We now have three models described by three formulas (Eqs. (8), (13) and (15)) for the persistence length of the AA wedge model, though only Eq. (15) takes into account all the features of the physical model. Calculations based on the three equations are presented in Fig. 3 as $P1$, $P2$ and $P3$. The results for Eqs. (8) and (13) differ by less than a part in a thousand and are not distinguishable on the graph. The dynamic factor α_d is used as a variable in the calculations and the parameters were set at the following values.

Fraction A (or T):	$p_A = 0.25$;	
AA/TT wedge angle:	$\theta_s = 8.7°$	[15];
helix angle:	$\theta_h = 360/10.4 = 34.6°$;	
AA bend at	$\phi_s = -155°$	[27];
TT bend at	$\phi_s = +155°$	[27].

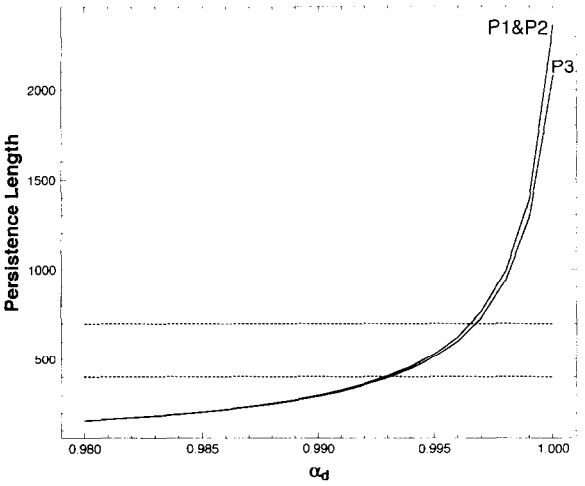


Fig. 3. Comparison of persistence lengths calculated by Eqs. (8), (13) and (15), referred to respectively as $P1$, $P2$ and $P3$. The horizontal lines represent the range in which persistence lengths are normally observed experimentally.

Eq. (15) was solved with both Mathematica and Fortran programs in terms of the unassigned variable α_d.

(a) Static persistence length. Angular correlations cause a distinct, but small, reduction in the persistence. At $\alpha_d = 1$ the dynamic contribution to the persistence disappears and $P = P_s$. Angular correlation reduces P_s from 2365 Å which is obtained with Eq. (8) or 12 to 2080 Å with Eq. (15), a reduction of about 7.5%. These numbers are dependent on the particular geometry of the AA wedge. Because of the screw axis between base pairs, conjugate sequences like AA and TT or AC and GT have related azimuths, ϕ_s, for the bend which they have in common. For example the AA bend of 8.7° is oriented at an angle of $-155°$ relative to the dyad axis using the angular convention of Bolshoy et al. A rotation of 180° about the interbase dyad axis converts this to the bend for the TT sequence at $+155°$. For convenience we define the angle Δ via the relation $\phi_s = 180° \pm \Delta$, where for AA and TT, $\Delta = 25°$ giving a 50° difference between the values of ϕ_s depending on whether it refers to TT or AA. If this angle is varied over the set $2\Delta = (0°, 45°, 90°, 135°$ and $180°)$, the static persistence length in Å assumes the set of values (2192, 2098, 1903, 1741, 1682). That this azimuthal effect can be quite large stems from the fact that at $2\Delta = 180$, AA sequences separated by 10 base pairs can act synergistically with TT base pairs at intermediate separations of 5 base pairs.

(b) Relaxed DNA. The horizontal lines in Fig. 3 represent rough upper and lower limits of the range of persistence lengths observed experimentally. The variation stems mostly from changes in ionic strength. For the dynamically bent segments an increase in α_d is a good representation of the ionic strength effect. For the wedge segments one would expect that at lower ionic strength θ_s would be decreased and the distribution leading to α_d would be distorted to favor straighter chains. There is no way that we can evaluate the latter effects at the present time. We will assume that the generic persistence length of B-form DNA is about 500 Å in the relaxed state at high ionic strength [33–35]. For a persistence length of 500 Å we interpolate α_d to be 0.9948, leading to $P_d = 654$ Å. Thus the presence of static AA bends reduces the persistence length from 654 to 500 Å. This is about a 25% contribution to the persistence length by static bends.*(c) Comparison of models.* The differences between models 1 and 3 is only of the order of 2–3% in the important region between the horizontal dashed lines of Fig. 3. Because of this an estimate of the contribution of static persistence length can be made on the basis of the wedge angle and the base composition alone, provided that the base composition is not too abnormal. It is the random positioning of static bends that lowers the persistence length. If wedges are highly correlated in sequence they result in curved contours for which the concept of persistence is not usually applicable. An extreme example is poly(dA:dT) which will have an infinite static persistence length.

(d) The additivity formula. Eq. (1) is astonishingly accurate for all three models. Using the relation $P_s = P(\alpha_d = 1)$, and evaluating P_d as a function of α_d with Eq. (3), we can compare values of $P1$, $P2$, and $P3$ with the calculated value from Eq. (1) over the full range of values of α_d. The relative error never exceeds 0.16% for $P3$ and 0.13% for $P1$ and $P2$. This shows that it is possible to do the single calculation of P_s using matrices $\langle \mathbf{M}_d \rangle$ and $\mathbf{M}_{IJ}$ with $\alpha_d = 1$, and then calculate $P3$ using Eqs. (1) and (3). Alternatively P_s could be estimated from a Monte Carlo calculation as was done by Trifonov et al. [15] and then use Eq. (3) to obtain $P3$ as a function of α_d. The value for P_s of 2160 Å obtained by them compares very favorably with the the analytical value of 2080 Å obtained from the $P3$ calculation. Finally one could do the approximate calculation mentioned in the previous paragraph and accept the small error. This clean separation of dynamic and static contributions probably stems from the fact that the model uses the same α_d for both straight and bent segments.

(e) Dependence on base composition. Fig. 4 shows the dependence of the persistence length on composition from the full range of $p_A = 0$, ($p_C = 0.5$; $p_G = 0.5$) to $p_A = 0.5$, $p_T = 0.5$. $P3$ represents the full calculation and $P1$, the approximation of Eq. (8). Only the central parts of the curves are meaningful since at the extrema there will be many homopolymeric stretches and alternating copolymers which deviate from the B-form. In the neighborhood of $p_A = 0.25$ both curves are nearly linear. It is seen that the AA wedge model predicts that DNAs which are 40% AT should have a persistence length which

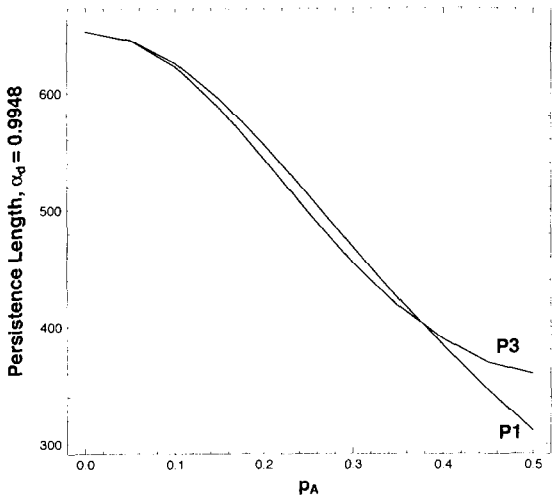


Fig. 4. Effect of variation of base composition on the persistence length of the AA wedge model using $\alpha_d = 0.9948$.

is about 20% longer than DNA's which are 60% AT. This is because this model has only one type of static bend. The effect is reduced when the small wedges of every dinucleotide step are considered, as will be seen in the next section.

The general wedge problem

The previous section has described a prototypical model: one with only one kind of wedge. Calculations on systems of this kind have been performed by a number of workers previously, but normally by computer methods involving molecular dynamics or Monte Carlo calculations rather than in closed form.

The work of the previous section can be generalized immediately to include a different type of wedge for each of the 16 possible pair-sequences. Such a calculation is now at least feasible: values for all 16 wedge angles and their orientations have been recently estimated, experimentally [27] and theoretically [14].

The generalization of the arguments above to four types of bases and sixteen types of pair-sequences leads to the 12 by 12 matrix

$$\mathbf{M}_{12} = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{T} \\ \mathrm{G} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{cccc} \mathrm{A} & \mathrm{T} & \mathrm{G} & \mathrm{C} \end{array} \\ \begin{pmatrix} p_A\mathbf{M}_{AA} & p_T\mathbf{M}_{AT} & p_G\mathbf{M}_{AG} & p_C\mathbf{M}_{AC} \\ p_A\mathbf{M}_{TA} & p_T\mathbf{M}_{TT} & p_G\mathbf{M}_{TG} & p_C\mathbf{M}_{TC} \\ p_A\mathbf{M}_{GA} & p_T\mathbf{M}_{GT} & p_G\mathbf{M}_{GG} & p_C\mathbf{M}_{GC} \\ p_A\mathbf{M}_{CA} & p_T\mathbf{M}_{CT} & p_G\mathbf{M}_{CG} & p_C\mathbf{M}_{CC} \end{pmatrix} \end{array} \tag{16}$$

Where p_J represents the probability of base J in the sequence and $\mathbf{M}_{IJ}$ is the transformation matrix for the sequence IJ. Since all pair sequences are potentially associated with static bends, the 3×3 submatrices have the form

$$\mathbf{M}_{IJ} = \mathbf{M}_h \cdot \begin{pmatrix} \frac{1+\alpha_d}{2}\left(c\theta_s \cdot c\varphi_s^2 + s\varphi_s^2\right) & \frac{1+\alpha_d}{2} c\varphi_s \cdot s\varphi_s (c\theta_s - 1) & \alpha_d \cdot c\varphi_s \cdot s\theta_s \\ \frac{1+\alpha_d}{2} c\varphi_s \cdot s\varphi_s (c\theta_s - 1) & \frac{1+\alpha_d}{2}\left(c\varphi_s^2 + s\varphi_s^2 \cdot c\theta_s\right) & \alpha_d \cdot s\varphi_s \cdot s\theta_s \\ -\frac{1+\alpha_d}{2} c\varphi_s \cdot s\theta_s & -\frac{1+\alpha_d}{2} s\varphi_s \cdot s\theta_s & \alpha_d \cdot c\theta_s \end{pmatrix}. \tag{B9}$$

The formula for the persistence length of the stochastic wedge model is then

$$P/\ell = \mathbf{lv}\left(\mathbf{I}_{12} + \sum_k^{\infty} \langle \mathbf{M}_{12} \rangle^k\right)\mathbf{rv}$$

$$= \mathbf{lv}\left(\frac{1}{\mathbf{I}_{12} - \langle \mathbf{M}_{12} \rangle}\right)\mathbf{rv} \tag{17}$$

where $\mathbf{I}_{12}$ is the 12×12 identity and $\mathbf{lv} = (0, 0, p_A, 0, 0, p_T, 0, 0, p_G, 0, 0, p_C)$, $\mathbf{rv} = (0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1)^t$.

Values for the thirty parameters (θ_s, φ_s and φ_h) required for the ten types of base sequences were obtained from [27] and [36]. These will be referred to as the BMHT parameters. This calculation must be regarded as highly provisional and is presented mainly to demonstrate the application of Eq. (17).

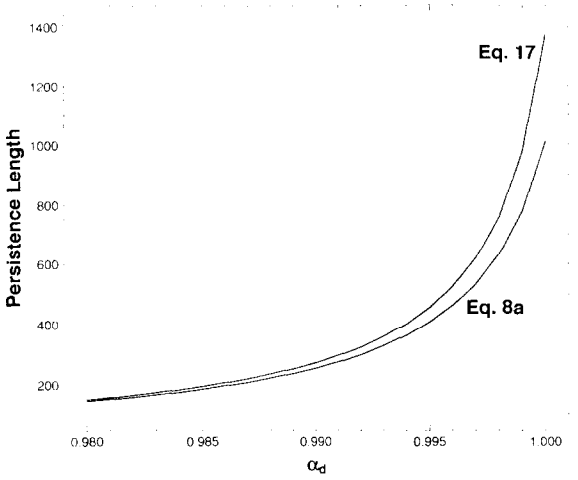


Fig. 5. Comparison of the complete multiwedge calculation (Eq. (17)) based on the parameters of Refs. [27] and [36] with a simple model using uncorrelated averages (Eq. (8a)).

The values of all 30 independent parameters are presumably at an initial stage of estimation. For example the change from the single-wedge to the many-wedge model caused θ_s for AA to change from 8.7 to 7.2 degrees. In addition the theoretical [14] and experimental values [27] are not the same.

The upper curve of Fig. 5 shows the persistence length of DNA calculated via Eq. (17). The curve is significantly lower than that obtained by means of the AA wedge model (Fig. 3) and extrapolates at $\alpha_d = 1$ to a static persistence length of 1370 Å. This is not surprising since this model has static bends, ranging from 0.8° to 7.6° at all but one of the base pair sequences. The AG = CT wedge has an even larger angle (7.6°) than AA = TT (7.2°). Taking this value of P_s and evaluating P_d as $(1-\alpha_d)^{-1}$, the additivity formula, Eq. (1), once again agrees with the results of the full matrix calculation to about one part in a thousand over the range of α_d of Fig. 5. If we assume that relaxed DNA has a total persistence length of about 500 Å, it is found via Eq. (1) that $P_d = 787$ Å and the fraction of $1/P$ which arises from static bends is about 36%. Even if the values of the parameters require further refinement, this may well be a appropriate estimate since the presence of small static bends is now being generally accepted. The interesting aspect of this calculation is that only about 2/3 of the loss of directional memory of DNA arises from flexibility, the other third being produced by the cumulative disorientation generated by randomly distributed small kinks.

The lower curve in Fig. 5 was calculated using Eq. (8a) in the form

$$P/\ell = \frac{1}{1-\langle\langle\cos\theta\rangle\rangle} = \frac{1}{1-\alpha_d\langle\cos\theta_s\rangle} \qquad (8b)$$

where $\langle\cos\theta_s\rangle$ is the average over all 16 bending angles of Ref. [27]. This calculation neglects both sequence and bending-orientation correlations, i.e, the relative orientation of two static bends in a chain depends on the number of base pairs between them. It is seen that the primitive model is not nearly so accurate when static bends are present in high density as it was in the single wedge model. The error in P_s is about 26%. For DNA in high salt with a total persistence length of 500 Å, the error is about 11%.

Application to specific DNA sequences

The generator matrices which have been used for the generic persistence length, i.e., averaged over all possible sequences, can be used for specific sequences as well with two changes: (1) the matrix powers must be replaced by an ordered product of matrices which follows the order of the sequence, (2) the 3×3 matrices, $\mathbf{M}_{IJ}$, given above in Eq. (B.9) are to be used rather than the 12×12 matrices which distribute sequences over all possibilities.

Assuming that the initial link of the chain is in the z direction, the link vector, $\boldsymbol{I}_m$, of the mth helical unit is given by

$$\boldsymbol{I}_m/\ell = \left(\prod_{k=2}^{m}\mathbf{M}_{I_{k-1}J_k}\right)\cdot\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}\langle x_m\rangle\\\langle y_m\rangle\\\langle z_m\rangle\end{pmatrix} \qquad (18)$$

where the index J_k is the value of J(= A,T,G, or C) associated with the kth unit in the helix and I_{k-1} is the value associated with the previous unit. The quantities $\langle x_m\rangle$, $\langle y_m\rangle$ and $\langle z_m\rangle$ are the average direction cosines of the mth link vector in the frame of the first link, and provide a measure of curvature and flexibility.

If the helix is rigid and straight the vector $\boldsymbol{I}_m/\ell$ will have the form $(0, 0, 1)^t$ meaning that the directional information of the first link is perfectly main-

tained and that the unit is rigid. (Superscript t means the transpose of a vector or matrix).

If the links are all straight on the average, but with thermal averaging over small distortions, then $\boldsymbol{I}_m/\ell = (0, 0, \langle z_m \rangle)^t$ with $\langle z_m \rangle < 1$. This indicates that the sense of the original direction is maintained but that it is not quantitatively transferred. The value of $\langle z_m \rangle$ is a real measure of flexibility (or rather of thermal flexing) with unity indicating perfect rigidity, $\langle z_m \rangle$ is the mth unit's contribution to the persistence length. For a simple chain governed by Eq. (3), $\langle z_m \rangle$ is α_d^{n-1}. Note that for this case most chains in the ensemble will have x and y components, but these will cancel out by symmetry.

If $\boldsymbol{I}_m/\ell$ has the form $(\langle x_m \rangle \langle y_m \rangle \langle z_m \rangle)^t$ with $\langle x_m \rangle$, $\langle y_m \rangle$ and $\langle z_m \rangle$ all different from zero, then the molecule is curved, since the mth link has an average direction different from the first. If $\langle x_m \rangle^2 + \langle y_m \rangle^2 + \langle z_m \rangle^2 = 1$, then the chain is perfectly rigid though bent or curved. This the case for pure static bends. The shape of the chain is deterministic and the averaging symbols can be dropped. If the initial unit in the chain is rotated by an angle, the entire chain including $\boldsymbol{I}_m$ is rotated. The x and y components are relative to the orientation of the initial unit of the DNA. The quantity $(x_m^2 + y_m^2)^{1/2}$ is a measure of the perpendicular component and $(x_m^2 + y_m^2)^{1/2}/z_m$ is the tangent of the angle between the first and mth links.

If $\boldsymbol{I}_m/\ell$ has the previous form but $\langle x_m \rangle^2 + \langle y_m \rangle^2 + \langle z_m \rangle^2 < 1$ then the chain is both curved and flexible. $(\langle x_m \rangle^2 + \langle y_m \rangle^2 + \langle z_m \rangle^2)^{1/2}$ is a measure of the determinacy of the chain contour over m links, with unity corresponding to perfect determinacy, and 0 a complete loss of determinacy because of dynamic flexibility. A rough estimate of the change in direction achieved at the mth unit may be obtained by evaluating $(\langle x_m \rangle^2 + \langle y_m \rangle^2)^{1/2}/\langle z_m \rangle$ as the tangent of an angle of bend. The angle calculated will be only a semi-quantitative representation, not a simply a mean angle, since one cannot do trigonometry on mean values of an ensemble of direction cosines.

Flexibility and curvature resulting from static bending are seen to be separable via Eq. (18) and the above criteria provide semi-quantitative measures for them.

The end-to-end vector, $\mathbf{h}$, of a DNA containing n base pairs can be generated by adding the contributions, $\boldsymbol{I}_m$ of the n links by vector addition,

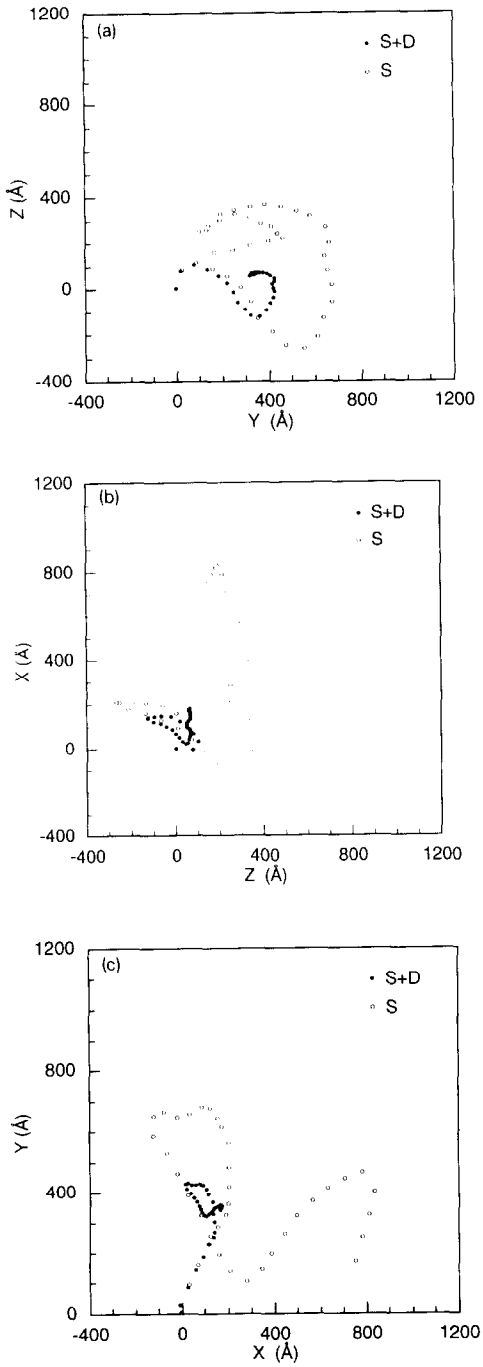
$$\mathbf{h} = \ell \left\{ \mathbf{I} + \sum \left(\prod_{k=2}^{m} \mathbf{M}_{I_{k-1}+J_k} \right) \right\} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \langle h_x \rangle \\ \langle h_y \rangle \\ \langle h_z \rangle \end{pmatrix}. \tag{19}$$

$\langle h_z \rangle$ is the persistence [16] of the chain and becomes the unperturbed persistence length [19] for $n \to \infty$; $\langle h_x \rangle$ and $\langle h_y \rangle$ are the mean distances that the chain has deviated from straightness when the nth link is reached; and $\boldsymbol{I}$ is the 3×3 identity matrix. Eqs. (18) and (19) combine to give a picture of the shape of a DNA molecule. Eq. (19) gives the mean x, y and z positions of the link and Eq. (18) calculates the mean projections of the link on the x, y and z-axes, thereby yielding information on both position and orientation. The value of $\langle z_m \rangle$ relative to unity indicates the decrease of directional memory owing to stochastic processes at the mth link. Comparing $|\mathbf{h}|$ with the contour length, L, indicates the loss of direction by both static and dynamic processes. In general these interpretations are available only by theoretical calculation, since the experimental quantities are the persistence length and the related properties $\mathbf{h}^2$ and the radius of gyration. On the other hand the statistical examination of electron micrographs can provide information that is a two-dimensional projection of the behavior described by Eqs. (18) and (19).

In Fig. 6 we present calculations based on these equations. The DNA molecule is a 976 base pair fragment of the kinoplast DNA of *B. caudatus* [37]. The BMHT parameters were used for static bending. Two calculations were performed: one (labelled D + S in the figure) with α_d set equal to 0.99568 to give a combined persistence length to the chain of 500 Å and a second (labelled S) with $\alpha_d = 1.0$ in order to visualize the static conformation. In the figure both chains start at the origin in the positive z direction.

The contour D + S represents the loci of average positions of the elements of the chain. Because of thermal agitation this chain loses a slight amount of directional information with each step from the starting link. As a result the link vectors connecting successive positions become shorter and shorter the further one is from the start. It is interesting to note

that the direction of each link of the averaged chain is identical to its direction in the static conformation. This is a result of the assumption of isotropy of thermal bending and is not a general characteristic of the chain-statistical model. The average chain mimics the features of the static chain but with ever diminishing scale, till, when the last (976th) unit is added the change in position of the endpoint is imperceptible.



The D + S curve can be compared with the simpler case of an ideal straight DNA with no static bends and a persistence length of 787 Å, which is the same as the dynamic persistence length of the D + S contour of Fig. 6. Because of symmetry all the link vectors would lie on the z-axis becoming shorter in a geometric progression with the limiting total length of 787 Å in complete analogy with the paradox of Zeno. It is clear that the 787 Å contour does not represent the contour of a single DNA molecule, but contains instead information on the average distance from the origin of each of the links of the chain.

Returning to Fig. 6 the D + S curve does not represent the contour of a DNA molecule, it is too short by a factor of 30. Calculation shows that at 976 base pairs the apparent length of a link has been reduced to 0.015 of its original length so that adding another 1000 units with the same thermal flexibility would only produce a small increment in the contour shown regardless of its intrinsic shape. By analogy with the statically straight helix, the far distant units of the chain will follow a Gaussian distribution about the limit point, which in this case is not on the z-axis.

This type of contour provides supplemental information to studies like that of Olson et al. [25] who have presented sample Monte Carlo calculations of the effect of dynamic fluctuations on the shapes of chains with curved equilibrium conformations. Their analysis provides a picture of the variety and spread of the contours of actual chains; the present analysis displays the average positions and gives a clear

Fig. 6. The average conformation of the kinoplast DNA of *B. caudatus* (976 bp) [37] The curves were generated with Eq. (19) using the Bolshoy et al. parameters. The contours marked S were obtained with $\alpha_d = 1$ and represent the molecule with static bends only. The contours marked D + S have $\alpha_d = 0.99568$. This value yields a total persistence length of 500 Å. Both chains start at the origin with the initial direction along the z-axis. They gradually separate as thermal agitation disorders the D + S contour. Points are shown at intervals of 25 bp, so that the convergence of the D + S curves to a single point are visible. (A) YZ projection; (B) ZX projection; (C) XY projection.

image of the apparent shrinkage of the chain because of dynamic flexibility.

3. Conclusions

This paper has presented an analytical development of the statistical propagation of DNA chains. In order to obtain formulas for the persistence length of a 'generic DNA', i.e., the average persistence length of all possible sequences of a given base composition, matrix methods were used to combine three features of DNA molecules which have sequence-dependent bends or 'wedges': (1) the geometrical features of the bends, i.e., their angular deviations and axes, (2) random variations of the direction of propagation resulting from local thermal fluctuations in structure, and (3) the proper distribution and correlation of the 16 pair sequences in a randomly generated sequence. A formula was also presented for application to specific finite sequences. The main general results are Eqs. (1), (15), (17), (18), and (19). Appendix A contains a demonstration that the persistence length of a heterogeneous DNA is an average over all persistence lengths starting from each link in the chain and going in both directions.

Any real DNA has a specified sequence, so it may differ from the generic DNA defined above. Since, as shown in Appendix A, the persistence length of a DNA is an average of all the persistence lengths which can be generated within it, a 'normal' DNA will be one which provides a representative sample, i.e., the same average persistence, as the total ensemble of all sequences. A curved DNA (or other abnormality) is one which, because of special sequences, generates a distribution which is not a good representation of the total ensemble.

Like any other current theory of DNA structure there are restrictive assumptions built into the model. In the present case there are three assumptions which need improvement. The first is that the thermal fluctuations around the equilibrium orientation obey an isotropic distribution. This is clearly an approximation since there is no structural isotropy in the DNA units. Both experiment [12,38] and theory [39,24,7] suggest that bends into the grooves would require less energy than bends about other axes. The second assumption is that DNA distortions can be understood in terms of nearest-neighbor models [14], i.e., that the local structure of DNA can be derived from interactions only involving two base pairs in sequence. This will be questioned by investigators of curved DNA, where, as discussed in a recent review [40], curvature has been definitely linked with the presence of tracts of A-sequences longer than two units. There is also substantial evidence that nearest neighbor models are not adequate for treating sequence-dependent twist, roll and tilt for general sequence DNA [41,24,42]. The third assumption is that fluctuations in the local helical angle have been ignored. Such fluctuations will affect the direction of the local curvature and hence the persistence properties of the chain. These assumptions have been made because of the lack of quantitative information to eliminate them: accurate potential functions for bending, and a specific geometry for A-tracts and their junctions with ordinary B-DNA. The matrix methods developed in the 50's and 60's are sufficiently powerful that they can be adapted to solve almost any problem in linear chain propagation, but in a field like molecular biophysics it is pointless to develop theories based on unknown parameters. These assumptions are under intensive investigation at the present time and will be discussed below.

The central part of the paper dealt with a prototypical model for the combination of static and dynamic bending mechanisms: the AA wedge model of Trifonov et al. [15]. This model was developed in three stages of increased refinement, partly to investigate the effect of various approximations, partly to compare with the results of other workers. Only the third stage leading to Eq. (17) contains all the aspects of the Trifonov model, but the formulas for the more primitive calculations turn out to be useful rough approximations. For example for the AA wedge model, in the region of experimental interest, Eq. (8) is in error by only a few percent compared to Eq. (17). This good agreement, however, depends on the orientation of the AA wedge, and becomes worse the further the axis or the wedge is removed from the local pseudo-dyad axis. The static persistence length, P_s, for the AA-wedge model contributes about 25% to the persistence length, the remainder coming from thermal fluctuations. These results are in favorable agreement with an earlier Monte Carlo calculation [15].

The results for the single wedge type were easily generalized to a formula, Eq. (17), in which all sixteen base pair sequences are associated with their own local parameters: bend angle, bend orientation and helical rotation. Within the assumptions mentioned above this is a complete solution of the problem, which averages over sequences, base pair distribution, and the geometric factors associated with the links of the helix. Using the BMHT parameters, calculations of the persistence length were made as a function of α_d and composition. For this model the contribution of static factors to the persistence length is about 33%.

This general theory was specialized to specific sequences and formulas were presented for the determination of the position and link orientation of the units of a DNA chain. These results are useful for following the contours of curved DNA's.

In a earlier paper [15] Trifonov et al. had conjectured that the reciprocal of the static and dynamic persistence lengths are simply additive to give the reciprocal of the net persistence length, Eq. (1). This formula was found to be very accurate (maximum error of the order of 0.1%) for all the formulas developed in this paper. As a result one only needs to calculate a static persistence length (putting $\alpha_d = 1$ in Eq. (15) or (17)) and add on a term for the dynamic persistence length using Eq. (1). There is a possibility that this good agreement depends on the assumption of isotropy, which gives a particularly simple form to Eqs. (5) and (6).

3.1. Comparison with other studies

While these studies were going on, there has been considerable activity in the field by other investigators. Very recent papers have appeared by the groups of Beveridge [30,21], Olson [25,22,23] and De Santis [14,43] as well as an earlier paper by Zhurkin et al. [20] Most of this work makes use of molecular dynamics, energy calculation and minimization, and Monte Carlo calculations. The calculations of energy and molecular dynamics go beyond the present work and will hopefully lead to an evaluation of the potential function for local conformational distortions of DNA so that the isotropic model can be supplanted. On the other hand the methodology of the present paper treats the stochastic part of the problem in an analytical way, which could obviate the need to Monte Carlo calculations on long chains. The method could be easily extended to anisotropic bending provided the local moments of the distribution were known from energy calculations.

One paper [25] comes particularly close to our own treatment in that a formula is derived which is equivalent to Eq. (8a). There is an important and interesting difference between the two approaches. We have followed Trifonov in considering a wedge to be a bend in the direction of the local helix axis. Olson et al. base their geometry on the roll and tilt angles of the plane of the base pair. This has one immediate advantage in that it automatically sets up an anisotropy in the bending distribution which favors bending into grooves. A disadvantage is that the simplicity of axial helical links is lost when the base planes are not perpendicular to the helix axis. Bending with roll (ρ) and tilt (τ) angles is discussed in Appendix B. It turns out that the two models are not exactly equivalent, though the differences are normally small. However, at the level of the primitive model (no correlation of the bend angles of adjacent links) which is discussed in the appendix of the cited paper, the formula which results is

$$P/\ell = \frac{1}{1 - \langle\langle \cos\rho\rangle\rangle\langle\langle \cos\tau\rangle\rangle}$$

This is derived both in the original reference and in Appendix B, though by different routes. The double brackets indicate an average over base pair sequences and over conformation. Ignoring compositional averages which contribute nothing to the argument we have the equivalence $\langle\cos\theta\rangle = \langle\cos\rho\rangle\langle\cos\tau\rangle$. For angles of the order of 10° the cosines can be expanded, retaining only the quadratic terms, to give the relations connecting the two descriptions of bends, $\theta_s^2 \cong \rho_0^2 + \tau_0^2$ and $\Delta\theta_d^2 \cong \Delta\rho^2 + \Delta\tau^2$ where the former are static angles and the latter are mean square deviations. See the discussion of Eq. (7).

Glossary

α, α_s, α_d $\langle\cos\theta\rangle$, $\cos\theta_s$, $\langle\cos\theta_d\rangle$

θ, θ_s, θ_d instantaneous bend angle of wedge, static wedge angle, instantaneous deviation from θ_s

φ_h	helix repeat angle
φ_s	orientational angle, relative to x-axis, of a static bend
σ, σ_s, σ_d	$\langle \sin\theta \rangle$, $\sin\theta_s$, $\langle \sin\theta_d \rangle$
C, S	$\cos\varphi_h$, $\sin\varphi_h$
h	rms end-to-end distance of DNA
$\boldsymbol{I}_m$	link vector of the mth link
L	contour length of DNA
ℓ	link length of DNA, ~3.4 Å
$\langle \mathbf{M}_d \rangle$	generator matrix for pure dynamic bending
$\mathbf{M}_h$	matrix for the helix operation
$\mathbf{M}_{IJ}$	generator matrix for a static + dynamic bend at step from base I to J in the double helix
p_s, p_d	probability of a static bend, dynamic (only) bend
p_J	probability of nucleotide J
P	total persistence length of DNA
P_s, P_d	static and dynamic persistence lengths
P_k^+, P_k^-	persistence lengths starting at unit k in forward and backward direction, respectively

Appendix A. Chain dimensions for systems with sequence-dependent persistence

By definition the mean square end-to-end distance is given by

$$h^2 = \langle \mathbf{h}\cdot\mathbf{h} \rangle = \left\langle \sum_k^n \boldsymbol{I}_k \cdot \sum_j^n \boldsymbol{I}_j \right\rangle = \sum_{k,j=i}^n \left\langle \boldsymbol{I}_k \cdot \boldsymbol{I}_j \right\rangle \quad \text{(A.1)}$$

where $\langle \cdots \rangle$ represents averaging over all conformations, h is the end-to-end vector, and $\boldsymbol{I}_k$ is the vector of the kth link. The persistence length is normally defined as

$$P_n = (1/\ell)\left\langle \boldsymbol{I}_i \cdot \sum_1^n \boldsymbol{I}_j \right\rangle \quad \text{(A.2)}$$

where n is large enough that the series converges. For a long inhomogeneous polymer, this is not very useful since it only deals with the persistence properties of one end of the molecule. We will want to consider persistence lengths that start on any unit, not only the first, and which may fail to converge when the starting point is too close to the end of a chain. We will also wish to consider persistence lengths which are defined in either direction. The notation P_k^+ and P_k^- will be used to denote forward and backward persistence lengths starting on the kth unit. With this notation (A.1) can be rearranged to

$$h^2 = \left\langle \sum_1^n \boldsymbol{I}_k \cdot \left(\sum_{j=k}^n \boldsymbol{I}_i + \sum_{j=k}^1 \boldsymbol{I}_j - \boldsymbol{I}_k \right) \right\rangle$$

$$= \ell \sum_k^n \left(P_k^+ + P_k^- - \ell \right). \quad \text{(A.3)}$$

In the last step it was assumed that all link-lengths are the same. For stiff chains like the worm-like chain the persistence lengths are long compared with the link-length so that ℓ can be ignored in the sum. If further the chain is homogeneous so that P does not depend on k except near the ends, and is also very long compared to P so that the ends can be ignored we have

$$h^2 = 2n\ell P_\infty = 2LP_\infty \quad \text{(A.4)}$$

where $L = n\ell =$ the contour length of the chain. This is a standard formula for the worm-like chain which relates chain dimensions to the persistence length. For a long heterogeneous chain with average link angles that depend on position, we can define a persistence length which is averaged over position and direction

$$P_{av} = 1/n \sum_1^n \frac{P_k^+ + P_k^-}{2} \quad \text{(A.5)}$$

Substituting in (A.3), ignoring ℓ compared to the persistence lengths,

$$h^2 = 2LP_{av} \quad \text{(A.6)}$$

Comparing this with (A.4) we see that for heterogeneous, stiff chains the usual persistence length is replaced by its average over position and direction.

Appendix B. Matrices and their averages

B.1. Isotropic bending of a straight segment

In the absence of known potentials simplified models must be used to evaluate the loss of persistence caused by thermal agitation. The simplest is isotropic flexing of one link in the helix relative to the next, giving an instantaneous change in helix

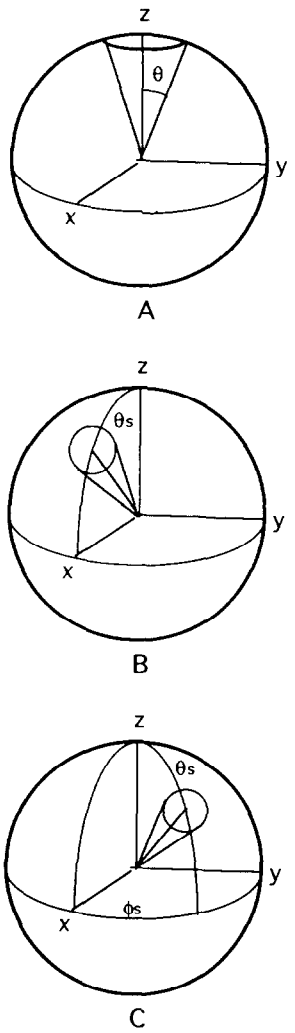


Fig. 7. Generation of isotropic distributions. (A) Axial distribution, represented by cone, generated by averaging the dynamic-bend matrix, Eq. (B.3); (B) the isotropic distribution at polar angles (θ_s, 0), Eq. (B.6); (C) the isotropic distribution at polar angles (θ_s, ϕ_s), Eq. (B.8).

direction (Fig. 1). Isotropic bending assumes a distribution function for the local orientation which depends on bending angle, but not on the orientation of the bend. The isotropic model will be assumed in the following. Fig. 7A shows one contour of such a distribution function about the polar axis. A bend inclined in the x direction is generated by rotating about the y-axis. The matrix for this rotation is

$$\mathbf{R}_y(\theta_d) = \begin{pmatrix} \cos\theta_d & 0 & \sin\theta_d \\ 0 & 1 & 0 \\ -\sin\theta_d & 0 & \cos\theta_d \end{pmatrix}. \qquad \text{(B.1)}$$

What is required for the analysis of the isotropic bending model is the matrix for bending about any axis in the equatorial plane, at the variable azimuth φ_d relative to the y-axis. The most expedient way to produce a bend at an arbitrary azimuth, φ_d, is to build it up by a series of rotations about the coordinate axes for which the matrices are simple and well known. The following scheme is the easiest: (1) Rotate the object about the z-axis till the desired bend direction is in the xz plane. This is done with matrix for a rotation by $-\varphi_d$ about the z-axis,

$$\mathbf{R}_z(-\varphi_d) = \begin{pmatrix} \cos\varphi_d & \sin\varphi_d & 0 \\ -\sin\varphi_d & \cos\varphi_d & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \text{(B.2)}$$

(2) The unit is now in a position where a bend about the y-axis is a bend in the φ_d direction, for which we utilize the matrix (B.1), (3) rotate the unit back into its original orientation about the z-axis with $\mathbf{R}_z(\varphi_d)$, the reverse of $\mathbf{R}_z(-\varphi_d)$. The matrix for a rotation by angle θ_d about an axis which is at angle φ_d relative to the y axis is then given as

$$\mathbf{M}(\theta_d) = \mathbf{R}_z(\varphi_d)\mathbf{R}_y(\theta_d)\mathbf{R}_z(-\varphi_d).$$

Written out in full this is

$$\begin{pmatrix} \cos^2\varphi_d\cos\theta_d + \sin^2\varphi_d & \cos\varphi_d\sin\varphi_d(\cos\theta_d - 1) & \cos\varphi_d\sin\theta_d \\ \cos\varphi_d\sin\varphi_d(\cos\theta_d - 1) & \cos^2\varphi_d + \sin^2\varphi_d\cos\theta_d & \sin\varphi_d\sin\theta_d \\ -\cos\varphi_d\sin\theta_d & -\sin\varphi_d\sin\theta_d & \cos\theta_d \end{pmatrix}.$$

This must be averaged over the dynamic bending angle θ_d and over the bending orientation specified by φ_d. Isotropic bending implies a uniform distribution over φ_d. As a consequence $\langle\cos\varphi_d\rangle = \langle\sin\varphi_d\rangle = \langle\cos\varphi_d\sin\varphi_d\rangle = 0$; $\langle\cos^2\varphi_d\rangle = \langle\sin^2\varphi_d\rangle = 1/2$. Also in our notation $\langle\cos\theta_d\rangle \equiv \alpha_d$. Substitut-

ing these values, we obtain the average bending matrix for isotropic, dynamic bending:

$$\langle \mathbf{M}(\theta_d) \rangle = \begin{pmatrix} (1+\alpha_d)/2 & 0 & 0 \\ 0 & 1+\alpha_d)/2 & 0 \\ 0 & 0 & \alpha_d \end{pmatrix}. \tag{B.3}$$

The (1, 1) and (2, 2) elements of this matrix measure the projection of the x- and y-axes of one unit onto the x- and y-axes of the next unit of the helix. For purely isotropic bending these projections do not enter into the problem, because the x and y projections of the link vector of the bending unit always vanishes by symmetry. In the presence of permanent wedges, however, the link vector is projected onto the x- and y-axes of an adjacent unit. These projections will be picked up and passed on by the isotropic units and can eventually be returned to the z direction by a second static bend and thus contribute to the persistence length. Mathematically the (1, 1) and (2, 2) components of the matrix are necessary for the projection of vectors which are not parallel to the z-axis.

It is to be expected that the x- and y-axes have different projection properties than the z-axis when the distribution is isotropic with respect to the z-axis. This can be visualized as follows. If θ_d is fixed and the bend orientation, φ_d, is varied, the local z-axis will trace out a polar circle as shown in Fig. 7A. At the same time the local x- and y-axes will trace out narrow figure eights about their original position. The average projection of the moving x- and y-axes on their original directions is $(1 + \cos \theta_d)/2$.

For helical DNA (B.3) must be multiplied by the matrix for the helical rotation

$$\mathbf{M}_n = \begin{pmatrix} C & -S & 0 \\ S & C & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{B.4}$$

where $S = \sin \phi_h$ and $C = \cos \phi_h$, to give the complete dynamic bending matrix

$$\langle \mathbf{M}_d \rangle = \begin{pmatrix} C(1+\alpha_d)/2 & -S(1+\alpha_d)/2 & 0 \\ S(1+\alpha_d)/2 & C(1+\alpha_d)/2 & 0 \\ 0 & 0 & \alpha_d \end{pmatrix}. \tag{B.5}$$

B.2. Static and dynamic bending for a bend at $\varphi = 0$

We consider now the presence of a static bend of fixed θ_s and $\phi_s = 0$. Superposed on this is thermal agitation about the fixed bend, again assumed to be isotropic. This distribution can be generated by rotating the preaveraged distribution of Fig. 7A by θ_s about the y-axis to produce Fig. 7B.

$$\begin{aligned} &\langle \mathbf{M}(\theta_s, 0) \rangle \\ &\quad = \mathbf{R}_y(\theta_s) \langle \mathbf{M}(\theta_d) \rangle \\ &\quad = \begin{pmatrix} \cos\theta_s(1+\alpha_d)/2 & 0 & \sin\theta_s\alpha_d \\ 0 & (1+\alpha_d)/2 & 0 \\ -\sin\theta_s(1+\alpha_d)/2 & 0 & \cos\theta_s\alpha_d \end{pmatrix}. \end{aligned} \tag{B.6}$$

Finally, this must be rotated by the helix matrix, (B.4), to give the complete matrix for a dynamically averaged, static bend at $\varphi = 0$,

$$\mathbf{M}_s(\theta_\gamma, 0) = \mathbf{M}_h \cdot \langle \mathbf{M}(\theta_\gamma, 0) \rangle. \tag{B.7}$$

This equation can be used for a simple model with only one type of static bend, by identifying the bend axis with the y-axis.

B.3. Static and dynamic bending for a bend at arbitrary azimuth

For this case we need a bend which forms an isotropic distribution about the orientation $(\theta_\gamma, \varphi_\gamma)$, Fig. 7C. This can be generated by the series of rotations

$$\langle \mathbf{M}(\theta_s, \varphi_s) \rangle = \mathbf{R}_z(\varphi_s) \langle \mathbf{M}(\theta_s, 0) \rangle \mathbf{R}_z(-\varphi_s) \tag{B.8}$$

where $\langle \mathbf{M}(\theta_s, 0) \rangle$ is given by Eq. (B.6). Written out this is

$$\langle \mathbf{M}(\theta_s, \varphi_s)\rangle = \begin{pmatrix} \frac{1+\alpha_d}{2}\left(c\theta_s \cdot c\varphi_s^2 + s\varphi_s^2\right) & \frac{1+\alpha_d}{2} c\varphi_s \cdot s\varphi_s (c\theta_s - 1) & \alpha_d \cdot c\varphi_s \cdot s\theta_s \\ \frac{1+\alpha_d}{2} c\varphi_s \cdot s\varphi_s (c\theta_s - 1) & \frac{1+\alpha_d}{2}\left(c\varphi_s^2 + s\varphi_s^2 \cdot c\theta_s\right) & \alpha_d \cdot s\varphi_s \cdot s\theta_s \\ -\frac{1+\alpha_d}{2} c\varphi_s \cdot s\theta_s & -\frac{1+\alpha_d}{2} s\varphi_s \cdot s\theta_s & \alpha_d \cdot c\theta_s \end{pmatrix} \tag{B.8}$$

To cut down the size of this formula for printing we have used the symbols c and s before angle variables to indicate sine and cosine. Before averaging, $\langle \mathbf{M}(\theta_s, \varphi_s)\rangle$ is a composite of six matrices, three for the dynamic bend and three for the static bend, which is reduced to four by averaging over φ_d. The complete matrix for the static, but thermally excited, bend of a unit is obtained by multiplying (B.7) by the matrix for the helical rotation, (B.4),

$$\mathbf{M}_{I,I} = \mathbf{M}_h \cdot \langle \mathbf{M}(\theta_s, \varphi_s)\rangle \tag{B.9}$$

There is a matrix for each of the sixteen sequential pairs of bases.

B.4. Bending in terms of tilt and roll angles

Another way to introduce bends in helical DNA is by successive rotations about axes fixed in a plane defined by the base pairs. These are the roll and tilt angles ρ and τ respectively. See Refs. [32] and [22] for precise definitions. With the proper choice of coordinates this can be considered as successive rotations about the x-axis, (τ-rotation) and the y-axis (ρ-rotation). The matrix for the bend followed by the helical operation is then,

$$\begin{aligned} \langle \mathbf{M}\rangle &= \langle \mathbf{R}_z(\phi_h) \cdot \mathbf{R}_y(\rho) \cdot \mathbf{R}_x(\tau)\rangle \\ &= \langle \mathbf{R}_z(\phi_h)\rangle \\ &\quad \cdot \begin{pmatrix} \cos\rho & \sin\rho\sin\tau & \sin\rho\cos\tau \\ 0 & \cos\tau & -\sin\tau \\ -\sin\rho & \cos\rho\sin\tau & \cos\rho\cos\tau \end{pmatrix}. \end{aligned} \tag{B.10}$$

This is Eq. (6) of [22]. It is of interest to compare the bending matrix of this model with (B.8), the bending matrix for a wedge. Euler's theorem states that these matrices are equivalent to a simple rotation about a single axis. We know the axis and rotation for Eq. (B.8): by definition the angle is θ_s and the axis is equatorial at azimuth ϕ_s. Euler angles like those of (B.10) are common in engineering and have been analyzed in Appendix B of [44]. It turns out that the two bends are not exactly equivalent. The axis of rotation of (B.10) is not equatorial but has a z component of $-\sin\rho/2 \sin\tau/2/\sin\theta'/2$, where θ' is the angle of the equivalent rotation and is given by $\cos\theta'/2 = \cos\rho/2 \cos\tau/2$. We note that for the small angles of DNA bends this gives the good approximation $\theta'^2 \cong \rho^2 + \tau^2$. For typical values of $\rho \approx 7°$ and $\tau \approx 2°$ [25], we find that $\theta' \approx 7.2°$ and the axis of rotation is 1° off the equatorial plane.

The difference is unimportant as long as one is comparing the isotropic versions of the two descriptions of DNA bending, since in this case it is only the zz elements of the matrices which contribute. Equating the zz elements of (B.8) and (B.10), assuming no fluctuations, we have $\cos\theta_s = \cos\rho_s \cos\tau_s$, which yields $\theta_s^2 \cong \rho_s^2 + \tau_s^2$, thereby justifying the equivalence of the descriptions for small angles. If one considers the angles to be fluctuating, then instead of the above, we have $\langle\theta^2\rangle \cong \langle\rho^2\rangle + \langle\tau^2\rangle$. Introducing static and fluctuating terms we obtain $\theta_s^2 \cong \rho_s^2 + \tau_s^2$, as before, and $\Delta\theta_d^2 \cong \Delta\rho^2 + \Delta\tau^2$, where the latter are mean square deviations (see Eq. (7)). We have assumed that ρ and τ are statistically independent.

Whether or not the non-equivalence of the two descriptions can be ignored for full matrix calculations will have to tested by detailed comparisons. The finite persistence length of DNA results from small angular deviations so caution will be required before the slight variation in axis direction can be considered of negligible importance.

References

[1] J.C. Marini, S.D. Levene, D.M. Crothers and P.T. Englund, Proc. Natl. Acad. Sci. USA, 79 (1982) 7664–7668.
[2] P.J. Hagerman, Ann. Rev. Biochem., 59 (1990) 755–781.

[3] R.E. Dickerson, in R.D. Wells and S.C. Harvey (Editors) Unusual DNA Structures, Springer Verlag, New York, 1988, pp. 287–306.
[4] E.N. Trifonov and J.L. Sussman, Proc. Natl. Acad. Sci. USA, 77 (1980) 3816–3820.
[5] H.-M. Wu and D.M. Crothers, Nature, 308 (1984) 509–519.
[6] E.N. Trifonov, CRC Crit. Rev. Biochem., 19 (1985) 89–106.
[7] A.R. Srinivasan, R. Torres, W. Clark and W.K. Olson, J. Biomol. Struct. Dyn., 5 (1987) 459–496.
[8] C.R. Calladine, H.R. Drew and M.J. McCall, J. Mol. Biol., 201 (1988) 127–137.
[9] R.C. Maroun and W.K. Olson, Biopolymers, 27 (1988) 561–564.
[10] M. Coll, C.A. Frederick, A.H.-J. Wang and A. Rich, Proc. Natl. Acad. Sci. USA, 84 (1987) 8385–8389.
[11] H.C.M. Nelson, J.T. Finch, B.F. Luisi and A. Klug, Nature, (1987) 221–226.
[12] K. Grzeskowiak, M. Kaczor-Grzeskowiak, D. Cascio and R.E. Dickerson, Biochemistry, 32 (1993) 8923–8931.
[13] D.S. Goodsell, M.L. Kopka, D. Cascio and R.E. Dickerson, Proc. Natl. Acad. Sci. USA, 90 (1993) 2930–2934.
[14] P. De Santis, A. Palleschl, M. Savino and A. Scipioni, Biochemistry, 29 (1990) 9269–9273.
[15] E.N. Trifonov, R.K.-Z. Tan and S.C. Harvey, in W.K. Olson, M.H. Sarma and M. Sundaralingam (Editors), DNA Bending and Curvature, Adenine Press. 1987.
[16] J.A. Schellman, Biopolymers, 13 (1974) 217–226.
[17] J.A. Schellman, Biophys. Chem., 11 (1980) 321–328.
[18] J.A. Schellman, Biophys. Chem., 11 (1980) 329–337.
[19] P.J. Flory, Statistical Mechanics of Chain Molecules, John Wiley, New York, 1969.
[20] V.B. Zhurkin, N.B. Ulyanov, A.A. Gorin and R.L. Jernigan, Proc. Natl. Acad. Sci. USA, 88 (1991) 7046–7050.
[21] C. Prevost, S. Louse-May, G. Ravishanker, R. Lavery and D.L. Beveridge, Biopolymers, 33 (1993) 335–350.
[22] N.L. Marky and W.K. Olson, Biopolymers, 34 (1994) 109–120.
[23] N.L. Marky and W.K. Olson, Biopolymers, 34 (1994) 121–142.
[24] C.S. Tung and S.C. Harvey, J. Biol. Chem., 261 (1986) 3700–3709.
[25] W.K. Olson, N.L. Marky, R.L. Jernigan and V.B. Zhurkin, J. Mol. Biol., 232 (1993) 530–554.
[26] M. Volkenstein, Configurational Statistics of Polymeric Chains, Interscience, New York, 1963.
[27] A. Bolshoy, P. McNamara, R.E. Harrington and E.N. Trifonov, Proc. Natl. Acad. Sci. USA, 88 (1991) 2312–2316.
[28] P. De Santis, S. Morosetti, A. Palleschi and M. Savino, in E. Clementi and S. Chin (Editors), Structure and Dynamics of Nucleic Acids, Proteins, and Membranes, Plenum, New York, 1986, pp. 31–49.
[29] V.B. Zhurkin, A.A. Gorin, A.A. Charakhchyan and N.B. Ulyanov, in D. Beveridge and R. Lowry (Editors), Theoretical Chemistry and Molecular Biophysics, Adenine Press, 1990, pp. 409–429.
[30] J.M. Withka, S. Swaminathon, J. Srinivasan, D.L. Beveridge and P.H. Bolton, Science, 255 (1992) 597–599.
[31] R.W. Wilson and J.A. Schellman, Biopolymers, 16 (1977) 2143–2165.
[32] R.E. Dickerson and E. Committee, EMBO J, 8 (1989) 1–4.
[33] P.J. Hagerman, Biopolymers, 20 (1981) 1503–1535.
[34] N. Borochov, H. Eisenberg and Z. Kam, Biopolymers, 20 (1981) 231–235.
[35] V. Rizzo and J. Schellman, Biopolymers, 20 (1981) 2143–2163.
[36] W. Kabsch, C. Sander and E.N. Trifonov, Nucleic Acids Res., 10 (1982) 1097–1105.
[37] J.D. Mulner, Z.A. Wood, T.J. Byers, K. Park, J. Griffith, S.L. Hajduk and S.C. Harvey, J. Mol. Biol., (in press).
[38] A.M. Burkhoff and T.D. Tullius, Cell, 48 (1987) 935–943.
[39] V.B. Zhurkin, Y.P. Lysov and V.I. Ivanov, Nucleic Acids Res., 6 (1979) 1081–1096.
[40] P.J. Hagerman, Biochim. Biophys. Acta, 1131 (1992) 125–132.
[41] R.E. Dickerson, J. Mol. Biol. 166 (1983) 419–441.
[42] J.R. Quintana, K. Grzeskowiak, K. Yanagi and R.E. Dickerson, J. Mol. Biol. 225 (1992) 379–395.
[43] P. De Santis, M. Fua, A. Palleschi and M. Savino, Biophys. Chem., 46 (1993) 193–204.
[44] H. Goldstein, Classical Mechanics, Addison-Wesley, Menlo Park, CA, 1980.