Polymer Chain Statistics and Conformational Analysis of DNA Molecules with Bends or Sections of Different Flexibility

Claudio Rivetti1, Chip Walker1 and Carlos Bustamante2*

1Institute of Molecular Biology
University of Oregon, Eugene
OR 97403, USA
2Howard Hughes Medical Institute and Institute of Molecular Biology, University of Oregon, Eugene, OR 97403 USA

The worm-like chain model has often been employed to describe the average conformation of long, intrinsically straight polymer molecules, including DNA. The present study extends the applicability of the worm-like chain model to polymers containing bends or sections of different flexibility. Several cases have been explicitly considered: (i) polymers with a single bend; (ii) polymers with multiple coplanar bends; (iii) polymers with two non-coplanar bends; and (iv) polymers comprised of sections with different persistence lengths. Expressions describing the average conformation of such polymers in terms of the mean-square end-to-end distance have been derived for each case. For cases (i) and (iv), expressions for the projection of the end-to-end vector onto the initial orientation of the chain are presented. The expressions derived here have been used to investigate DNA molecules with sequence-induced bending (A-tracts). Mean-square end-to-end distance values determined from a large number of A-tract containing DNA molecules visualized by scanning force microscopy resulted in an average bend angle of 13.5° per A-tract. A similar study was performed to characterize the flexibility of double-stranded DNA molecules containing a single-stranded region. Analysis of their mean-square end-to-end distance yielded a persistence length of 1.3 nm for single-stranded DNA.

Keywords: worm-like chain; persistence length; scanning force microscopy; A-tracts; ssDNA

Introduction

Two models are often used to describe the entropic elasticity of long polymer molecules, the freely jointed chain (FJC; Flory, 1969) and the worm-like chain (WLC; Kratky & Porod, 1949; Schellman, 1974; Landau & Lifshitz, 1980, 1986). The first model considers the polymer as a series of orientationally independent statistical segments (Kuhn segments), whose length is a direct measure of the chain stiffness. The WLC model, on the other hand, treats the polymer as a relatively stiff rod made up of a homogeneous elastic material (elastica). In this model, the stiffness of the chain is described by its persistence length, i.e. the distance over which the memory of the initial orientation of the polymer persists.

Recently, single-molecule manipulation methods utilizing magnetic beads and optical tweezers (Smith et al., 1992, 1996; Perkins et al., 1994; Baumann et al., 1997; Wang et al., 1997) have been applied to investigate the elastic properties of DNA molecules throughout a full range of forces and extensions. These studies have exposed the limitations of the FJC model as a description of the elastic behavior of DNA, and have established instead the validity of the WLC model. Statistical chain parameters derived from the WLC model, such as the mean-square end-to-end distance and the mean directional cosine of the molecule, have also been successfully applied to the conformational analysis of intrinsically straight DNA molecules visualized by electron microscopy (EM; Muzard et al., 1990; Bednar et al.,...
considerable effort has also been invested in the study of DNA molecules containing bends or regions of different flexibility. Previous studies have shown that intrinsically curved DNA is often associated with various cell processes, from gene regulation to chromatin packaging (Hagerman, 1990; van der Vliet & Verrijzer, 1993; Perez-Martin et al., 1994). Similarly, regions of DNA with different elastic properties arise during processes such as replication, recombination and DNA repair, where DNA molecules consisting of both double-stranded and single-stranded regions are present as intermediates of the reaction.

First observed in kinetoplast DNA (Simpson, 1979; Challberg & Englund, 1980), runs of four or more adenine residues (A-tracts) introduce intrinsic dichroism and NMR (for reviews, see Hagerman, 1992; Dickerson et al., 1996). DNA curvature has also been studied by high-resolution microscopy (EM and SFM), where the fraction of bent versus straight conformations, (Griffith et al., 1986) or the end-to-end distance distribution of DNA molecules (Hansma et al., 1994) has been used to describe the DNA bending. The increased flexibility of DNA molecules containing nicks, single-stranded gaps, or bulged bases has also been characterized by EM (Wang et al., 1992; Le Cam et al., 1994). However, the analysis of these results is somewhat limited by the lack of a quantitative theory directly relating the average DNA conformation to the bend angle or to the presence of regions with increased flexibility.

The aim of this work is to present a theoretical description of several statistical parameters for polymers with static bends or consisting of sections with different elastic properties, within the framework of the WLC model. The derived expressions have been used to quantify sequence-induced DNA bending from the mean-square end-to-end distance of a large number of DNA molecules visualized by SFM. Similarly, the mean-square end-to-end distance of DNA molecules harboring a single-stranded gap has been used to determine the persistence length of the single-stranded DNA region.

**Theory**

The first part of this section summarizes the main results of the WLC model (Kratky & Porod, 1949; Schellman, 1974; Landau & Lifshitz, 1980, 1986) for isotropic, intrinsically straight, homogeneous polymers. In the second part, these expressions will be generalized to include polymers containing static bends or consisting of sections with different flexibility.

**Intrinsically straight polymers**

In the WLC or persistent chain model, the polymer is treated as a straight, relatively stiff rod made up of a continuous, homogeneous, isotropic material. The elastic behavior of the chain is assumed to be purely entropic, i.e. at finite temperatures, thermal fluctuations induce local curvature in the polymer making it bend and deviate smoothly from the straight configuration.

An important result of the WLC model is that the average directional correlation between two segments in a polymer decreases exponentially with their separation:

\[
\langle \cos(\theta) \rangle = \langle \vec{u}(s) \cdot \vec{u}(s') \rangle = e^{-|s-s'|/P}
\]

where \(\vec{u}(s)\) and \(\vec{u}(s')\) are the unit vectors tangent to the chain at positions \(s\) and \(s'\), respectively. \(P\) is the persistence length of the chain, i.e. the decay length through which the memory of the initial orientation of the polymer persists. The persistence length is a statistical mechanical quantity whose value depends both on the mechanical properties of the chain through its bending rigidity, \(k\), and on the magnitude of the statistical deformations induced in the polymer by the energy of the thermal bath: \(P = \kappa/k_B T\) (Landau & Lifshitz, 1980).

Equation (1) shows that the directional correlation among polymer segments in the WLC is a multiplicative function of their distance along the chain. Thus, the directional correlation between segments at positions \(s\) and \(s'\) can be written as:

\[
\langle \vec{u}(s) \cdot \vec{u}(s') \rangle = \langle \vec{u}(s) \cdot \vec{u}(\ell) \rangle \langle \vec{u}(\ell) \cdot \vec{u}(s') \rangle
\]

where \(\vec{u}(\ell)\) is the unit tangent to the chain at any position \(\ell\) intermediate between \(s\) and \(s'\). This result will play an important role in the derivations that follow.

Equations (1) and (2) can be used to obtain a number of statistically relevant parameters of the chain. For example, from equation (1) it is possible to derive the average projection of the end-to-end vector \(\vec{R}\) on the initial unit tangent of the chain, i.e.:

\[
\langle \vec{u}(0) \cdot \vec{R} \rangle = \int_0^L \langle \vec{u}(0) \cdot \vec{u}(s') \rangle ds' = \int_0^L e^{-|s'-s|/P} ds' = P(1 - e^{-L/P})
\]

where \(L\) is the contour length of the chain and \(P\) its persistence length. This projection attains the value of \(P\) when \(L \gg P\), and therefore can be used as an operational definition of persistence length. Also,
for very rigid, or for very short molecules, \( P \gg L \),
the projection approaches the value \( L \).

Another important statistical parameter of the
WLC is the mean-square end-to-end distance,
which can be written as (Landau & Lifshitz,
1980):

\[
\langle R^2 \rangle = \int_0^L ds \int_0^L (\tilde{u}(s) \cdot \tilde{u}(s')) ds' = \int_0^L ds \int_0^L e^{-\vert s - s' \vert / P} ds' \tag{4}
\]

Upon integration this yields:

\[
\langle R^2 \rangle = 2PL \left( 1 - \frac{P}{L} \right) (1 - e^{-L/P}) \tag{5}
\]

where for \( L \gg P \), \( \langle R^2 \rangle \rightarrow 2PL \) and for \( P \gg L \),
\( \langle R^2 \rangle \rightarrow L^2 \). In the intermediate regime where
\( L > P \), the approximate formula:

\[
\langle R^2 \rangle \approx 2PL \left( 1 - \frac{P}{L} \right) \tag{6}
\]

may be used with less than 5% error when
\( L \gg 3P \).

\textbf{Intrinsically bent polymers}

A static bend located at any position along a
polymer reduces its global dimensions. Let a
single bend \( \beta \) occur at a distance \( \ell \) from one end
of a polymer with contour length \( L \) (Figure 1a).
It is convenient to consider the polymer as consisting of
three sections extending from \( s \) to \( \ell - \delta / 2 \),
from \( \ell - \delta / 2 \) to \( \ell + \delta / 2 \), and from
\( \ell + \delta / 2 \) to \( s' \), respectively, where \( \delta \) is a number
that can be chosen arbitrarily small. Then, using the
multiplicative property of the correlation function,
equation (2), it is possible to define the orientational correlation between segments
located on both sides of the bend:

\[
\langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle_\beta = \langle \tilde{u}(s) \cdot \tilde{u}(\ell - \delta / 2) \rangle_\beta \langle \tilde{u}(\ell + \delta / 2) \rangle_\beta \langle \tilde{u}(s') \rangle_\beta \tag{7}
\]

In the limit of \( \delta \rightarrow 0 \), \( \langle \tilde{u}(\ell - \delta / 2) \cdot \tilde{u}(\ell + \delta / 2) \rangle_\beta =
\cos \beta \), and equation (7) becomes (see the derivation in the Appendix):

\[
\langle \cos(0) \rangle_\beta = \cos \beta \langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle_\beta = \cos \beta e^{-|s-s'|/P} \tag{8}
\]

This result is independent of the specific location of the bend angle between positions \( s \) and \( s' \)
in the polymer and represents a generalization of equation (1).

Equation (8) can now be used to derive the corresponding generalizations of equations (3) and (4)
for bent polymers. For a chain with a bend \( \beta \) at position \( \ell \), the average projection of the end-to-end vector on the initial unit tangent is:

\[
\langle \tilde{u}(0) \cdot \tilde{R}_\beta \rangle = \int_0^\ell \langle \tilde{u}(s) \rangle_\beta \tilde{u}(0) \cdot \tilde{u}(s') ds' = \int_0^\ell \langle \tilde{u}(s) \rangle_\beta \tilde{u}(0) \cdot \tilde{u}(s') ds' \tag{9}
\]

Similarly, the mean-square end-to-end distance of the bent chain can be written as:

\[
\langle R^2 \rangle_\beta = \int_0^\ell ds' \int_0^\ell \langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle_\beta ds + 2 \int_0^\ell ds' \int_0^\ell \langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle_\beta ds' \tag{10}
\]

where \( s \) and \( s' \) denote locations on either side of the bend located at position \( \ell \) from one end,
and as before, \( \tilde{u}(s) \) and \( \tilde{u}(s') \) are the unit vectors
tangent to the chain at position \( s \) and \( s' \), respectively. Equation (10) shows that the mean-square end-to-end distance of a bent polymer can be partitioned into three contributions: The first and
the third terms of the right-hand side represent the orientational correlation between segments located
on the same side of the bend. These terms have the same form as equation (4),
expressing the fact that the polymer sections at both sides of the bend behave as intrinsically
straight worm-like chains. The second term represents the correlation between segments located
on opposite sides of the bend, and its value depends on the magnitude of the bend. Using the
multiplicative property of the correlation function, the integrand in this term can be written
as in equation (8) so that equation (10) becomes:

\[
\langle R^2 \rangle_\beta = \int_0^\ell ds' \int_0^\ell e^{-|s-s'|/P} ds + 2 \int_0^\ell ds' \int_0^\ell \cos \beta e^{-|s-s'|/P} ds
\]

Performing the integration yields:

\[
\langle R^2 \rangle_\beta = 2PL \left( \frac{P}{L} - (1 - e^{-\ell \beta}) \right) \tag{11}
\]

For bend angles up to 90°, an approximate form of equation (12) may be utilized if both \( \ell \) and
Figure 1. A representation of bent and non-homogeneous worm-like chains and behavior of their mean-square end-to-end distances. All graphs are in arbitrary length units. a, A polymer chain with contour length $L$ including a static bend angle $\beta$ at position $\ell$ along the contour. $\vec{R}$ is the end-to-end vector; $\vec{u}(s)$ and $\vec{u}(s')$ are the unit vectors tangent to the chain at a location $s$ and $s'$ on opposite side of the bend. $\beta$ is the end-to-end vector; $\vec{u}(s)$ and $\vec{u}(s')$ are the unit vectors tangent to the chain at a location $s$ and $s'$ on opposite side of the bend. b, $(R^2)$ determined by equation (12) as a function of the bend angle $\beta$ located in the center of a chain with unit contour length and persistence length. c, Same as in b but in this case the bend was fixed to 90° and its position was varied from one end of the polymer to the other. d, A polymer chain with contour length $L$ and persistence length $P$, containing two non-coplanar bends $\beta_1$ and $\beta_2$. $\ell_1$, $\ell_2$ and $\ell_3$ are the contour lengths of the three sections such that $\ell_1 + \ell_2 + \ell_3 = L$. $\psi$ is the phase angle between the $\beta_1$ and $\beta_2$ planes. $\vec{R}$ is the end-to-end vector; $\vec{u}_1$, $\vec{u}_2$ and $\vec{u}_3$ represent the initial orientation of the first, second and third sections, respectively; $\vec{u}(s)$ and $\vec{u}(s')$ are the unit vectors tangent to the chain at the locations $s$ and $s'$ in non-adjacent polymer sections. e, Three-dimensional representation of the relative orientations of unit vectors tangent to the polymer shown in d. The orthogonal system has been chosen in such a way that $\vec{u}_2$ lies on the $z$ axis. For convenience, the drawing in e has been rotated 90° with respect to the drawing in d. $\psi$ is the polar angle and $\phi$ the phase angle. Notice that when $\psi = 0$, i.e. $\phi = 90^\circ$, all vectors are in the $z$-$y$ plane. f, $(R^2)$ determined by equation (18) as a function of the phase angle $\psi$. For the calculation, $\beta_1$ and $\beta_2$ were set to 90°, $P$ was set to 1 and the section contour lengths were: $\ell_1 = 0.4$, $\ell_2 = 0.2$ and $\ell_3 = 0.4$. g, A polymer chain of contour length $L$ consisting of two sections of length $\ell$ and $L - \ell$, and persistence lengths $P_1$ and $P_2$ respectively. $\vec{R}$ is the end-to-end vector; $\vec{u}(s)$ and $\vec{u}(s')$ are the unit vectors tangent to the chain at a location $s$ and $s'$ in different polymer sections. h, $(R^2)$ determined by equation (22) as a function of the persistence length $P_2$ of the second section. Both sections were assumed to have the same contour length ($L = 1$; $\ell = 0.5$) and persistence length $P_1 = 1$. i, Same as in h but in this case the length of the first section ($\ell$) was varied from 0 to 1. The persistence length $P_1$ and $P_2$ were 1 and 0.1, respectively.
L - \ell is greater than \(2P\). If these conditions are met, then:

\[
\langle R^2_{\beta_i} \rangle \approx 2PL \left( 1 - \frac{P}{L} (2 - \cos \beta) \right) \tag{13}
\]

This equation can be used with no more than 5% error. Larger bend angles can be accommodated by equation (13) to similar accuracy if \(L - \ell > 3P\).

Notice that equation (12) is symmetric with respect to the location of the bend, and reduces to equation (5) for \(\beta = 0^\circ\). Furthermore, just as equation (5) reduces to the geometric value of \(L^2\) when \(P > L\), equation (12) reduces to the law of cosines in the same limit. This expression also shows that the mean-square end-to-end distance of a bent polymer, \(\langle R^2 \rangle\), decreases as the bend angle increases over the range from 0 to 180° (Figure 1b), and reaches a minimum when the bend angle is located in the middle of the polymer (Figure 1c).

Equation (12) can be readily generalized for the case of a chain of length \(L\) containing \(N\) coplanar bends separating \(N + 1\) polymer sections of length \(\ell_n\) with \(n = 1, 2, \ldots, N + 1\). Bends are said to be coplanar if they occur in the same plane at 0°K, i.e. in the absence of thermally induced fluctuations:

\[
\langle R^2_{\beta_1, \ldots, \beta_N} \rangle = 2PL \left\{ 1 - \frac{P}{L} \left[ \sum_{n=1}^{N+1} (1 - e^{-\ell_n/P}) \right. \right.

\left. - \sum_{n=1}^{N} \cos \beta_n (1 - e^{-\ell_n/P})(1 - e^{-\ell_{n+1}/P}) \right.

\left. - \sum_{n=1}^{N-1} \sum_{m=n+2}^{N+1} \cos \left( \sum_{j=n}^{m-1} \beta_j \right) \left( \prod_{j=n+1}^{m-1} e^{-\ell_j/P} \right) \right.

\left. \times \left( 1 - e^{-\ell_m/P}(1 - e^{-\ell_{m+1}/P}) \right) \right\} \tag{14}
\]

where \(\ell_1\) is the distance along the polymer from one end to the first bend, \(\ell_{N+1}\) is the distance from the last bend to the opposite end of the polymer, and \(\ell_n\) is the distance between the \(n\)th and \((n - 1)\)th consecutive bends. Equation (14) shows that the mean-square end-to-end distance of a polymer with multiple coplanar bends can be separated into three contributions: the first summation arises from the orientational correlation between tangent vectors within the same section. The second summation results from the correlation between vectors located in adjacent sections, and the double summation from correlations between vectors located in non-adjacent sections. If every \(\ell_i\) \((i = 1 \text{ to } N + 1)\) is greater than \(2P\), then the contribution from non-adjacent sections is negligible and the approximate formula:

\[
\langle R^2_{\beta_1, \ldots, \beta_N} \rangle \approx 2PL \left( 1 - \frac{P}{L} \left( N - \sum_{n=1}^{N} \cos \beta_n \right) \right) \tag{15}
\]

may be utilized with less than 5% error.

**Polymers with two non-coplanar bends**

It is also of interest to derive an expression for the mean-square end-to-end distance of a polymer containing two non-coplanar bends. As shown in Figure 1d, the two bends \(\beta_1\) and \(\beta_2\) divide the polymer into three sections of contour length \(\ell_1, \ell_2,\) and \(\ell_3\) such that \(\ell_1 + \ell_2 + \ell_3 = L\). The mean-square end-to-end distance of the polymer can then be written as:

\[
\langle R^2_{\beta_1, \beta_2} \rangle = \int_0^{\ell_1} ds \left( \langle \ddot{u}(s) \cdot \ddot{u}(s') \rangle ds \right.

\left. + \int_0^{\ell_2} ds \left( \langle \ddot{u}(s) \cdot \ddot{u}(s') \rangle ds - 2 \int_0^{\ell_1} ds \langle \ddot{u}(s) \cdot \ddot{u}(s') \rangle ds \right) \right.

\left. + \int_0^{\ell_3} ds \langle \ddot{u}(s) \cdot \ddot{u}(s') \rangle ds \right)

\right. + 2 \int_0^{\ell_1} ds \langle \ddot{u}(s) \cdot \ddot{u}(s') \rangle ds \left. + 2 \int_0^{\ell_2} ds \langle \ddot{u}(s) \cdot \ddot{u}(s') \rangle ds \right. \tag{16}
\]

The first three terms describe the orientational correlations between segments located in the same section of the polymer and are the same as those obtained for a polymer containing a single bend. The fourth and fifth terms account for the correlations between segments located in adjacent sections of the polymer on each side of the bends and are identical with those obtained for the case of a polymer with a single bend. The last term corresponds to correlations between segments located in non-adjacent polymer sections.

The magnitude of the correlation between any two segments \(\ddot{u}(s)\) and \(\ddot{u}(s')\) located in non-adjacent sections is obtained as follows. First, \(\ddot{u}(s)\) is made to coincide with the unit tangent at the beginning of the first segment \(\ddot{u}_1\), and \(\ddot{u}(s')\) is made to coincide with the unit tangent at the beginning of the third section, \(\ddot{u}_3\). Next, the geometric relationship among the three unit tangents \((\ddot{u}_1 = u(s), \ddot{u}_2,\) and \(\ddot{u}_3 = u(s'))\) is determined at the location of \(\ddot{u}_2\) at \(\ell_1\). This is done by reducing the length of the \(\ell_1\) and \(\ell_2\) sections. This relationship can be seen in Figure 1e, where the tangent \(\ddot{u}_2\) has been made to coincide with the \(z\)-axis of the laboratory frame, \(\ddot{u}_1\) is a vector on the \(y-z\) plane forming an angle \(\beta_1\) with \(\ddot{u}_2\), and \(\ddot{u}_3\) is a vector forming an angle \(\beta_2\) with \(\ddot{u}_2\) and whose projection on the \(x-y\) plane...
makes the angle $\phi$ with the x-axis. For convenience, a phase angle $\psi$ is defined as the complement of $\phi$ so that $\beta_1$ and $\beta_2$ are coplanar when $\psi = 0$. Using this construction, it can be shown that:

$$\tilde{u}(s) \cdot \tilde{u}(s') = \cos \beta_1 \cos \beta_2 - \sin \beta_1 \sin \beta_2 \cos \psi \quad (17)$$

Finally, $\tilde{u}_1$, $\tilde{u}(s)$, $\tilde{u}_2$, and $\tilde{u}(s')$ are returned to their original locations on the polymer, allowing for the loss of memory of their relative orientations to take place. Using the multiplicative property of the orientational correlation function, and the geometric relationship in equation (17), the mean-square end-to-end distance of the chain can be written as:

$$\langle R^2_{\tilde{p}_1, \tilde{p}_2} \rangle = 2PL \left[ 1 - \frac{P}{L} \left( 1 - e^{-\xi_1/P} \right) + \left( 1 - e^{-\xi_2/P} \right) - \cos \beta_1 (1 - e^{-\xi_1/P})(1 - e^{-\xi_2/P}) - \cos \beta_2 (1 - e^{-\xi_1/P})(1 - e^{-\xi_2/P}) - (\cos \beta_1 \cos \beta_2 - \sin \beta_1 \sin \beta_2 \cos \psi) \times e^{-\xi_1/P} (1 - e^{-\xi_1/P})(1 - e^{-\xi_2/P}) \right] \quad (18)$$

When $\psi$ is zero, the trigonometric factor of the last term becomes $\cos(\beta_1 + \beta_2)$ and equation (18) reduces to the case of a polymer with two coplanar bends. As shown in Figure 1f, the mean-square end-to-end distance reaches a minimum at $\psi = 0^\circ$ and attains a maximum at $\psi = 180^\circ$.

**Polymers with sections of different flexibility**

From the multiplicative property of the orientational correlation function it is possible to obtain the statistical parameters of WLCs harboring sections of different flexibility. Figure 1g shows a polymer chain of contour length $L$, consisting of two sections of lengths $\ell$ and $L - \ell$, and persistence lengths $P_1$ and $P_2$, respectively. The orientational correlation between two tangent vectors located in the same polymer section is again given by equation (1). In analogy with equation (2), the orientational correlation between any two tangent vectors in different polymer sections can be written as:

$$\langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle = \langle \tilde{u}(s) \cdot \tilde{u}(\ell) \rangle \langle \tilde{u}(\ell) \cdot \tilde{u}(s') \rangle$$

$$e^{-|s' - s|/P_{app}} = e^{-|s' - s|/P_1} \times e^{-|s' - s|/P_2} \quad (19)$$

where $P_{app}$ is the apparent persistence length of the entire polymer. Taking $s' = L$ and $s = 0$, the last expression leads to the relation $L/P_{app} = \ell/P_1 + (L - \ell)/P_2$, an expression that embodies the law by which the contour length and the persistence length of each section combine to yield the apparent persistence length of the overall polymer.

Unlike the case of a uniform chain, the apparent persistence length of a blocked polymer is an intensive property of the material and depends on the relative size of the individual sections. However, caution must be exercised in the interpretation of $P_{app}$, as this quantity is valid only for tangents taken at specific sites on the molecule ($s = 0$, $s' = L$ and $\ell$ in this case). That is, measurements of tangents at these three positions over a large number of molecules to obtain $\langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle$ would find that the autocorrelation function between $s = 0$ and $s' = L$ behaves as if the intervening polymer had an apparent persistence length given by the above relation. Thus, $P_{app}$ cannot be used in equation (5) to determine $\langle R^2 \rangle$, since it correlates only the behavior of three segments in the chain and not all possible correlations between $s$ and $s'$, as implied by the integrals in equation (5).

The presence of sections with distinct elastic properties affects the statistical parameters of the chain. In particular, the average projection of the end-to-end vector onto the initial tangent of the polymer can be obtained from:

$$\langle \tilde{u}(0) \cdot \tilde{R} \rangle_{P_1, P_2} = \int_0^L \int_0^L e^{\xi_1/P_1} ds' + \int_0^L \int_0^L e^{\xi_2/P_2} ds'$$

$$= P_1 (1 - e^{-\xi_1/P_1}) + P_2 (1 - e^{-\xi_2/P_2}) \quad (20)$$

As in the case of a polymer with one bend, the mean-square end-to-end distance of a two-section polymer can be partitioned into three contributions:

$$\langle R^2_{\tilde{p}_1, \tilde{p}_2} \rangle = \int_0^L \int_0^L \langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle ds' ds + 2 \int_0^L \int_0^\ell \langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle ds' ds$$

$$+ \int_0^L \int_0^L \langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle ds' ds \quad (21)$$

Using equation (19) and integrating yields:

$$\langle R^2_{\tilde{p}_1, \tilde{p}_2} \rangle = 2P_1 \ell \left( 1 - \frac{P_1}{L} (1 - e^{-\xi_1/P_1}) \right) + 2P_2 (L - \ell)$$

$$\times \left( 1 - \frac{P_2}{L - \ell} (1 - e^{-\xi_2/P_2}) \right)$$

$$+ 2P_1 P_2 (1 - e^{-\xi_1/P_1})(1 - e^{-\xi_2/P_2}) \quad (22)$$

Equation (22) reduces to equation (5) when $P_1 = P_2$. The mean-square end-to-end distance of a two-section polymer as a function of their relative sizes and persistence lengths is plotted in Figure 1h and i, respectively. This expression may be further simplified in the case where both $\ell > 2P_1$ and $L - \ell > 2P_2$. Under these conditions an approximate formula is obtained:
\[
\langle R_{P_1,P_2}^2 \rangle \approx 2 \left( P_1 \ell \left( 1 - \frac{P_1}{L} \right) + P_2(L - \ell) \left( 1 - \frac{P_2}{L - \ell} \right) + P_1P_2 \right)
\]  

Equation (23) can be used with less than 5% error.

Equations (12) and (22) can be generalized for the case of a polymer that contains \( N \) coplanar bends of different magnitude at the junctures between \( N + 1 \) polymer sections that may have different persistence lengths. The mean-square end-to-end distance of such a polymer can be shown to be:

\[
\langle R_{\beta_1,...,\beta_n;P_1,...,P_N}^2 \rangle = 2 \left\{ \sum_{n=1}^{N+1} P_n \ell_n \left[ 1 - \frac{P_n}{\ell_n}(1 - e^{-\beta_n/P_n}) \right] + \sum_{n=1}^{N} P_nP_{n+1} \cos \beta_n \times (1 - e^{-\beta_n/P_n})(1 - e^{-\beta_n/P_n}) + \sum_{n=1}^{N-1} \sum_{m=n+2}^{N} P_nP_m \cos \left( \sum_{j=n+1}^{m-1} \beta_j \right) \left( \prod_{j=n+1}^{m-1} e^{-\beta_j/P_j} \right) (1 - e^{-\beta_m/P_m}) \right\}
\]  

Equation (24) can accommodate polymers with multiple sections of different persistence length but no bends, by simply setting each \( \beta = 0^\circ \). Moreover, if all sections have equal persistence length, equation (24) reduces to equation (14).

All the equations derived above are valid for polymers in three dimensions. However, with the exception of equation (18), these expressions can describe polymers in two dimensions if the persistence length is multiplied by a factor of 2 to account for the loss of one degree of freedom in the molecule (Frontali et al., 1979). Since the WLC model applies to polymers at thermodynamic equilibrium, the equations derived here are applicable only to the case of molecules deposited in two dimensions that have equilibrated prior to the determination of their average dimensions.

### Results

A conformational analysis of double-stranded DNA containing bends or an intervening single-stranded region, provides a direct application of the equations presented above. In particular, the mean-square end-to-end distance, \( \langle R^2 \rangle \), will be used first to quantify the bend angle of DNA molecules possessing a number of phased A-tracts. Second, the persistence length of a single-stranded DNA segment inserted between two double-stranded fragments will be determined from the \( \langle R^2 \rangle \) of the entire polymer. In both cases, the end-to-end distance distributions are obtained from a large number of DNA molecules imaged in air by SFM. The molecules are deposited onto freshly cleaved mica under conditions that permit their equilibration among accessible configurations in two dimensions (Bustamante & Rivetti, 1996; Rivetti et al., 1996), thus making the study of these polymers amenable to treatment by the two-dimensional analogs of the above expressions. In each case the full expressions derived above have been used to ensure that the interpretation of the results does not depend on the ratio of polymer contour length to its persistence length.

### DNA molecules containing phased A-tracts

DNA fragments containing two to eight phased A-tracts, previously described by (Thompson & Landy, 1988), were imaged by SFM to determine the magnitude of their bends. Each A-tract consists of six adenine residues followed by a four base-pair G+C-rich spacer. The insertion of each additional A-tract module near the center of a DNA fragment increases both the bend angle and the contour length of the molecule. Figure 2a to d shows four selected images of fragments containing two, four, six and eight A-tracts, respectively. The trend towards more highly bent molecules as the number of A-tracts in the molecule increases is evident. From a large set of images collected in different experiments, the contour length and end-to-end distance of each individual molecule were measured as described in Materials and Methods; the average values are reported in Table 1. An experimental rise per base-pair of 3.07 Å was determined from the SFM images by dividing the measured average contour length by the total number of base-pairs.

Approximating the DNA curvature as a single bend could represent an oversimplification of the DNA conformation, since the overall curvature of these DNA fragments is due to a series of phased A-tracts, each introducing a small bend. For example, in the fragments containing eight A-tracts, the bends are distributed over an 80 bp region within a 443 bp molecule. Therefore, the average conformation of these molecules is best described by equation (14), which accounts for polymers containing multiple coplanar bends.

For calculation purposes, the location of the \( N \) bends that divide the polymer into \( N + 1 \) sections were assumed to occur at the beginning of each A-tract. This choice will not affect the results, since the measurements are not sensitive enough to distinguish placement of the bend at the edge or in the center of the adenine residues, or even in the intervening G+C-rich sequences. Thus, in each of the seven cases the length from one end of the polymer to the first A-tract, \( L_1 \), is 156 bp, whereas the length from the beginning of the last A-tract to
the end of the polymer, $\ell_{N+1}$, is 217 bp. Each of the $N-1$ intervening sections are 10 bp long. The contour length of each section was obtained by using the experimental raise of 3.07 Å/bp. The persistence of all the sections was assumed to be 53 nm (Bustamante et al., 1994; Rivetti et al., 1996).

Although the expressions derived in Theory assume that the bend angles are rigid and not subject to thermal fluctuations, this assumption will not affect the $\langle R^2 \rangle$ of relatively long molecules (~400 bp), since the variance in the end-to-end distance is dominated by the flexibility of the DNA on both sides of the bend. This effect has been verified by simulating DNA molecules in which bends of 13.5° were either fixed or had the natural DNA flexibility. Comparison of $\langle R^2 \rangle$ obtained from ensembles of 5000 molecules in each case showed no significant difference between the two types of bending (data not shown).

Figure (3) shows the mean-square end-to-end distance versus the number of A-tracts for each DNA fragment (filled circles). The error bars indicate the deviation from the mean of three independent experiments. The continuous line represents the least-squares fit of equation (14) to the data obtained with a bend angle $\beta = 13.5^\circ$ per A-tract. The dotted lines are the $\langle R^2 \rangle$ determined with equation (14) using, from top to bottom, $\beta$ values

Table 1. Contour length and mean-square end-to-end distance values for the series of DNA fragments with increasing number of A-tracts

<table>
<thead>
<tr>
<th>Number of A-tracts</th>
<th>Length (bp)</th>
<th>Contour length (nm)</th>
<th>$\langle R^2 \rangle$ (nm²)</th>
<th>Number of molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>383</td>
<td>117.7</td>
<td>9161</td>
<td>618</td>
</tr>
<tr>
<td>3</td>
<td>393</td>
<td>121.6</td>
<td>9414</td>
<td>657</td>
</tr>
<tr>
<td>4</td>
<td>403</td>
<td>125.5</td>
<td>9531</td>
<td>671</td>
</tr>
<tr>
<td>5</td>
<td>413</td>
<td>127.7</td>
<td>8780</td>
<td>751</td>
</tr>
<tr>
<td>6</td>
<td>423</td>
<td>129.7</td>
<td>7741</td>
<td>772</td>
</tr>
<tr>
<td>7</td>
<td>433</td>
<td>131.8</td>
<td>7436</td>
<td>577</td>
</tr>
<tr>
<td>8</td>
<td>443</td>
<td>133.8</td>
<td>6952</td>
<td>610</td>
</tr>
</tbody>
</table>
of 10, 12, 15, and 17°. The curves make a maximum because, as each A-tract is inserted into the DNA fragment, the mean-square end-to-end distance tends to decrease due to increased bending, but it tends to increase because the molecule also becomes 10 bp longer. Initially, when only two or three A-tracks are present, the modest amount of DNA bending is not sufficient to counter the effect of the increased contour length. It is only when five or more A-tracks are present that the reduction in \( \langle R^2 \rangle \) becomes evident.

**DNA molecules with a single-stranded gap**

Double-stranded DNA fragments harboring a single-stranded gap were imaged by SFM and the persistence length of the single-stranded region was determined from the mean-square end-to-end distance of the entire polymer. Since ssDNA is more flexible than dsDNA, it is expected that this section will behave much like a flexible hinge, thereby decreasing the mean-square end-to-end distance of the polymer. Moreover, the strength of this effect should depend on the persistence length and on the length of the single-stranded gap.

Several DNA constructs were made (see Material and Methods) in which two dsDNA fragments were connected by single-stranded oligonucleotides of various lengths. Using this procedure, gaps of 1, 3, 5, 10, 15 and 20 thymine residues were introduced between double-stranded fragments 299 and 337 bp long. These constructs can therefore be considered as polymers comprised of three sections where the persistence length of the first and third sections is the same, but different from the persistence length of the intervening single-stranded region. Thymine was chosen for the composition of the single-stranded region because of the weak stacking interactions between these bases (Saenger, 1984).

Figure 4a to e depicts SFM images of DNA fragments with no gap, along with those of fragments with one, three, five and ten residue gaps. These images show that molecules with a single-stranded gap can assume a V-shaped conformation, which becomes more pronounced as the gap size increases. In some molecules the gap is not revealed by the tip, whereas in others it appears as a globular feature or a depression near the center of the DNA fragment (Figure 4f and g). The contour length and the end-to-end distance of the molecules were measured from the images as described in Material and Methods; the average values for each DNA fragment are shown in Table 2. An experimental rise of 2.90 Å/bp was determined using the average contour length of all the DNA fragments with a single-stranded region. It was assumed that the single-stranded region did not contribute to the measured contour length of the molecule. This choice appears justified, since the contour length of fragments with a single-stranded gap was slightly smaller than that of the fully double-stranded molecule. Compare, for example, the contour length of the DNA fragments with zero and five residue gaps listed in Table 2. It is possible that the presence of the highly flexible single-stranded region allows overlapping of the two connected dsDNA fragments, thereby reducing the measured contour length. This could also explain the presence of a globular feature at the gap site. Thus, for the calculations that follow, the contour length of the two double-stranded regions is determined using the experimental rise, whereas the contour length of the single-stranded gap is obtained using a rise per nucleotide of 7 Å (Saenger, 1984). The persistence length of the dsDNA was assumed again to be 53 nm (Bustamante et al., 1994; Rivetti et al., 1996).

Figure 5 shows the mean-square end-to-end distance as a function of the number of thymine residues in the single-stranded gap (filled circles). Also plotted is the \( \langle R^2 \rangle \) as given by equation (24) using the following parameters: \( \ell_1 = 86.7 \text{ nm} \), \( \ell_2 = 0.7 \times \).
(the number of single-stranded nucleotides) nm, $\ell_3 = 97.7$ nm and $P_1 = P_3 = 53$ nm. In this case the cosine terms of equation (24) are all equal to 1 ($\beta_i = 0$) due to the absence of intrinsic bend angles at the junctions between segments. The lines represent the behavior of equation (24) for different persistence length values of the single-stranded region. Starting from the lower dotted line, $P_2$ equals 0.8 nm, 1.0 nm, 1.3 nm, 1.6 nm, 2.0 nm, 2.4 nm and 2.8 nm. As seen in the Figure, equation (24) with $P_2 = 1.3$ nm (continuous line) agrees reasonably well with the data points corresponding to DNA fragments with single-stranded gaps up to five thymine residues. For longer gaps, the measured mean-square end-to-end distance is larger than the predicted values (see Discussion).

**Discussion**

**Comparison with other studies**

The theory presented above is applicable to polymers that can be modeled as homogeneous, isotropic chains. At the base-pair level, DNA is neither homogenous nor isotropic, however. Recent studies that have modeled DNA as a heterogeneous, anisotropic polymer go beyond the present work with respect to the conformational

<table>
<thead>
<tr>
<th>Gap size (number of T)</th>
<th>Length (bases)</th>
<th>Contour length (nm)</th>
<th>$&lt;R^2&gt;$ (nm$^2$)</th>
<th>Number of molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (no gap)</td>
<td>641</td>
<td>187.4</td>
<td>20,204</td>
<td>708</td>
</tr>
<tr>
<td>1</td>
<td>656</td>
<td>195.3</td>
<td>19,309</td>
<td>954</td>
</tr>
<tr>
<td>3</td>
<td>639</td>
<td>189.0</td>
<td>16,787</td>
<td>667</td>
</tr>
<tr>
<td>5</td>
<td>641</td>
<td>184.4</td>
<td>15,486</td>
<td>687</td>
</tr>
<tr>
<td>10</td>
<td>646</td>
<td>179.5</td>
<td>15,197</td>
<td>667</td>
</tr>
<tr>
<td>15</td>
<td>651</td>
<td>184.8</td>
<td>15,696</td>
<td>519</td>
</tr>
<tr>
<td>20</td>
<td>656</td>
<td>181.3</td>
<td>15,555</td>
<td>736</td>
</tr>
</tbody>
</table>
details of the molecule (Olson et al., 1993; Schellman & Harvey, 1995). On the other hand, this work arrives at analytical expressions that describe the statistical parameters of the DNA molecule in two and three dimensions, thus reducing the need for computer-intensive molecular simulations. Moreover, anisotropic effects are likely to be more relevant at the level of local protein-DNA interactions but should affect only weakly the global statistical parameters of the double helix.

**Experimental sensitivity**

The expressions presented here for bend angle determination are valid for any contour length, bend angle and persistence length. However, the statistical deviations in $\langle R^2 \rangle$ resulting from the natural flexibility of the polymer limit their effective range of applicability. The experimental sensitivity of the method depends then on the ability to determine a change in $\langle R^2 \rangle$ over and above the natural variability of this quantity. One measure of the sensitivity of this approach is the fractional change in $\langle R^2 \rangle$ due to the bend, i.e. $s = 1 - \langle R^2 \rangle / \langle R^2 \rangle$. As expected, the sensitivity reaches a maximum when the bend is located in the middle of the polymer (not shown), and increases with the magnitude of the bend angle (Figure 6a). On the other hand, factors that tend to increase the number of conformations available to the chain at a given temperature will decrease the sensitivity of the measurement. Increasing the contour length of the polymer (Figure 6b) or decreasing the persistence length, for example, will lower the experimental sensitivity (Figure 6a and b). Because the persistence length of a polymer equilibrated in two dimensions is twice that of the same polymer equilibrated in three dimensions, SFM data are particularly convenient for bend angle determination from the mean-square end-to-end distance measurements. A similar argument holds for the case of blocked polymers, where the sensitivity of the $\langle R^2 \rangle$
Measurements of A-tract DNA bending

The value of 13.5° per A-tract determined from the mean-square end-to-end distance analysis presented here is in the lower end of bend angles reported for A-tract-containing DNA molecules. Indeed, previous studies have reported a wide range of values, from 11° to 45° per A-tract, depending upon the technique used to characterize them, the specific A-tract sequence, the salt concentration and the temperature (Crothers et al., 1990; Chan et al., 1993; Olson & Zhurkin, 1996). While it is safe to rule out mica-DNA interactions as a possible source of bias (Rivetti et al., 1996), other factors could, in principle, influence the bending measured with SFM. For example, the deposition buffer used here contains Mg²⁺, which has been shown in gel migration studies to increase the DNA curvature and/or stabilize A-tracts (Diekmann, 1987; Bruckner et al., 1994). According to these studies, however, Mg²⁺ concentrations as used in the deposition buffer employed here (2 mM) have only a small effect on gel mobility, and higher (~10 mM) concentrations are needed to produce optimal effects. On the other hand, it is known that A-tract bending exhibits a rapid monotonic decrease as temperature is raised from 5 to 60°C when the spacer sequences between A-tracks are G+C-rich (Diekmann, 1987). Thus, the somewhat lower values for A-tract bending obtained here could simply reflect the fact that all depositions were carried out at room temperature. Attempts to corroborate the temperature/bending relationship were unsuccessful, however, as depositions carried out at 4°C yielded too few DNA molecules on the mica surface to support a statistical analysis. Further studies along the lines shown here may eventually clarify the interplay of temperature and Mg²⁺ concentration on the magnitude of A-tract bending.

Finally, it cannot be ruled out that the process of placing a three-dimensional molecule onto a two-dimensional surface may result in somewhat reduced values of bending per A-tract. Long DNA fragments made up of repeated, in-phase A-tract-containing elements may assume a superhelical configuration and be therefore inherently non-planar (Calladine et al., 1988, 1991). Estimates for DNA containing repeated A-tract elements similar to those used here suggest that to form a complete turn of the superhelix requires around 190 to 200 base-pairs of DNA, or 19 to 20 repeat units. Thus, for the largest number of repeat units used in this study (eight), such a molecule would be expected to form about half a turn of a superhelix. This could introduce a significant non-planar distortion of the DNA axis, which could, in turn, bias the deposition towards molecules possessing smaller bending angles. If this were the case, the distortion and any resulting bias should increase with the number of repeat elements in the DNA fragment and the amount of bending per A-tract should reflect this increasing bias. However, the data shown in Figure 3 fall in a narrow range of bending per A-tract and do not display any clear trend away from a value of 13.5° per A-tract as the number of repeat units is increased.

Significantly, not all A-tract-containing fragments appear bent in the images. This observation has also been reported in EM studies of kinetoplast DNA (Griffith et al., 1986) and has been used to support the hypothesis that the A-tract region may interconvert between two discrete bent and straight conformations (Hagerman, 1990). However, this observation might be a consequence of the natural flexibility of the bent region, which may or may not be the same as that of the surrounding sequences: the flexibility of the bent region itself may lead to a locally straight molecular conformation. Alternatively, an intrinsic bend may be undetectable sometimes if a nearby bend in the opposite direction, due to normal DNA fluctuations, occurs close enough so that the two bends cannot be resolved independently by the microscope. Again, such a molecule would appear straight in the image.

A test was done to investigate whether or not microscope observations of straight molecules can be unequivocally interpreted as resulting from the interconversion of the bent regions between two discrete (bent and straight) configurations. To this end, the fraction of straight molecules found in SFM images of DNA fragments containing six, seven and eight A-tracts was compared to the fraction of straight molecules observed in computer-generated worm-like chains containing the same number of bends. The computer-generated molecules were visualized as simulated SFM images displaying the current spatial resolution of the microscope. To also investigate the effects of flexibility, two kinds of bent molecules were simulated, one with a fixed, rigid bend of 13.5° representing each A-tract, and the other with the same bend but possessing the normal DNA flexibility (see Materials and Methods). The case where the bend interconverts between two discrete configurations was not simulated. Two criteria were used to determine the number of straight molecules. The first consisted of estimating the DNA bending by simple inspection; a molecule was considered straight if the angle formed by the two vectors tangent to the DNA contour on either side of the A-tract region appeared less than ~45°. This choice of angle is arbitrary, but it serves as a convenient point of separation between DNA molecules that are clearly bent and those that appear straight within the region of interest. The second criterion consisted of the analysis of the correlation between two vectors representing the natural flexibility of the bent region, which may or may not be the same as that of the surrounding sequences: the flexibility of the bent region itself may lead to a locally straight molecular conformation. Alternatively, an intrinsic bend may be undetectable sometimes if a nearby bend in the opposite direction, due to normal DNA fluctuations, occurs close enough so that the two bends cannot be resolved independently by the microscope. Again, such a molecule would appear straight in the image.
Table 3 shows that, as expected, the probability of seeing a straight molecule decreases as the number of A-tracts increases. More importantly, no significant difference in the percentage of straight configurations is observed among the real A-tract molecules and the simulated polymers possessing either a fixed bend or a naturally flexible bend, regardless of the method employed for comparison. These observations show that, to the level of resolution of the microscope, the fraction of straight configurations in the experimental images is equally well explained by the two alternative models described above. Therefore, that interpretation of microscopic data as evidence of the interconversion of A-tracts between discrete configurations based on the molecular appearance alone is not justified.

Table 3 shows that, as expected, the probability of seeing a straight molecule decreases as the number of A-tracts increases. More importantly, no significant difference in the percentage of straight configurations is observed among the real A-tract molecules and the simulated polymers possessing either a fixed bend or a naturally flexible bend, regardless of the method employed for comparison. These observations show that, to the level of resolution of the microscope, the fraction of straight configurations in the experimental images is equally well explained by the two alternative models described above. Therefore, that interpretation of microscopic data as evidence of the interconversion of A-tracts between discrete configurations based on the molecular appearance alone is not justified.

Because of many experimental similarities, the methods described here can be applied to determine the magnitude of protein-induced DNA bending by high-resolution microscopy. Because the bending is localized at the protein-binding site, bend angle estimates can be obtained from equation (12) as applied for two dimensions. Currently, bend angle measurements of protein-DNA complexes visualized by SFM are done by drawing the tangents to the entry and exit points of the DNA on both sides of the bound protein. Although direct, the tangent method has several limitations. First, it is often difficult to decide the exact location of the entry and exit points of the DNA from the protein and to determine precisely the tangents to the DNA at these points. In fact, the choice of these tangents is a subjective process that easily accounts for the high variability of bend angle determination among different operators analyzing the same data (unpublished results).

Table 3. Percentage of straight DNA molecules observed in SFM and simulated images of A-tract containing DNA fragments

<table>
<thead>
<tr>
<th>Number of A-tracts</th>
<th>SFM (eye method) (%)</th>
<th>Simulation fixed angle (eye method) (%)</th>
<th>Simulation flexible angle (eye method) (%)</th>
<th>SFM (tangents) (%)</th>
<th>Simulation fixed angle (tangents) (%)</th>
<th>Simulation flexible angle (tangents) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>18</td>
<td>19</td>
<td>17</td>
<td>22</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>18</td>
<td>13</td>
<td>16</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>13</td>
<td>10</td>
<td>12</td>
<td>10</td>
<td>11</td>
</tr>
</tbody>
</table>

**sSDNA persistence length determination**

It has also been shown here that a statistical analysis of DNA molecules with single-stranded gaps can be used to determine the persistence length of the ssDNA. The exponential decay of \( R^2 \) as the gap size increases (Figure 5), in contrast with the sinusoidal behavior of the \( R^2 \) as a function of the bend angle (Figure 1b), confirms that single-stranded gaps are regions of increased flexibility and not of static bending.

Various determinations of the persistence length of ssDNA exist in the literature. A value of 0.75 nm was determined in 150 mM salt by pulling general sequence ssDNA using laser tweezers techniques (Smith et al., 1996), while a range of values ranging from 1.05 nm in 150 mM salt to 4.0 nm in 10 mM salt were obtained in 8 M urea using fluorescence recovery after photobleaching (Tinland et al., 1997). While the current study was performed at low (10 mM) monovalent salt concentration, the buffer also contained 2 mM Mg\(^{2+}\), which causes a significant decrease in the persistence length of ssDNA as determined by laser tweezers measurements (Martin Hegner, personal communication). Furthermore, the single-stranded region in the present study is comprised entirely of thymine residues, which are known to have minimal stacking interactions and that may also contribute to a somewhat more flexible chain (Saenger, 1984). Thus, the value of 1.3 nm obtained here is consistent with the estimates arrived at by these other methods.

The measured \( R^2 \) reaches a constant value after a gap of five single-stranded thymine residues is introduced into the DNA template. In contrast, a PAGE analysis of DNA comprising several 23 nt long double-stranded regions separated by single-stranded gaps (Mills et al., 1994) indicated that the flexibility introduced by the single-stranded gaps saturates after only two single-stranded residues. Calculations using the two-dimensional form of equation (24) (data not shown) revealed instead that the flexibility of a DNA molecule containing a single gap should saturate somewhat faster than the flexibility of a similar molecule with multiple gaps. Therefore, it is possible that this discrepancy reflects not the difference in placement of the single-stranded regions, but the limited sensitivity of gel electrophoresis to subtle changes in molecular conformations.
When the single-stranded gap is longer than five nucleotides, the observed \(<R^2>\) is larger than the theoretical prediction. Several factors may contribute to the more extended conformations observed for gaps larger than five nucleotides, such as excluded volume, the formation of secondary structures within the ssDNA region, or a differential interaction of ssDNA and dsDNA with the mica surface. In a similar study (Rivetti et al., 1996), it has been shown that excluded volume interactions can increase the mean-square end-to-end distance of DNA molecules when the contour length is above 20 persistence lengths. Although the DNA fragments used here are shorter than 20 persistence lengths, the presence of a flexible region in the middle of the molecule increases the probability that the two arms will come into close proximity, and can allow self-avoiding interactions to become evident at shorter contour lengths. Moreover, this effect should be enhanced for molecules in two dimensions. Finally, equation (22) should be useful to investigate the effect of the flexibility of loops, bulges and other common RNA structural elements on the statistical dimension of these molecules.

Conclusions

This study has presented a new application of the worm-like chain model to polymers with intrinsic bends or sections of different persistence length. The equations derived here are completely general, and apply to any polymer that can be adequately described by the WLC model. In particular, the mean-square end-to-end distance has been shown to be a useful quantity in describing the average conformation of such polymers. Moreover, the expressions derived above seem especially well suited for use in conjunction with high-resolution microscopy techniques, such as SFM or EM, due to the gain in sensitivity obtained with molecules deposited on a surface. Determination of the magnitude of DNA bends from the mean-square end-to-end distance can also become useful in the determination of protein-induced bending. This method circumvents the need to estimate the tendgts to the DNA molecule at both sides of the protein, a process that leads to the highest variability of the scoring of the bend angle among different operators.

Materials and Methods

Sample preparation

DNA with phased A tracts

DNA fragments containing from two to eight phased A-tracts were generated by PCR amplification using pJT170-2 through pJT170-8 (Thompson & Landy, 1988) as templates. The DNA primers 5'-CTAGCGCTA-TATGCGTTGATG-3' and 5'-CAGCACGCCATAGTGAC-TGG-3' (Biosource, Menlo Park, CA) were used in each reaction. PCR was performed under standard reaction conditions using Pfu DNA polymerase (Stratagene). The PCR products were analyzed by agarose gel electrophoresis. A single band with the expected mobility was observed in all cases. The DNA was purified using a QIAquick PCR purification kit (Qiagen). The elution step was performed with 40 μl of 10 mM Tris-HCl (pH 8.5) and followed by phenol/chloroform extraction. Samples with a concentration of about 200 nM DNA were stored at −20°C.

DNA with a single-stranded gap

The strategy adopted to construct DNA fragments with single-stranded gaps of 3, 5, 10, 15 and 20 nt consisted of the ligation of two dsDNA fragments by means of a single-stranded oligonucleotide as shown in Figure 7a. The oligonucleotide sequence was such that 4 bp of the 3' and 5' ends were complementary to the 3' and 5' overhangs of the 299 and 337 bp DNA fragments, respectively. A number of thymine residues in the middle of the bridging oligonucleotide, referred to here as SSNT, determined the size of the single-stranded gap. For better yield, the ligation was carried out in two steps. A DNA fragment 720 bp long, containing unique NspI and XhoI restriction sites, was obtained by PCR amplification of the region 2038 to 2757 of pDE13 plasmid using Deep Vent DNA polymerase and 5'OH primers under standard reaction conditions. After amplification the DNA was phenol/chloroform-extracted, ethanol-purified and resuspended in TE buffer. The DNA was digested overnight with XhoI endonuclease in the recommended reaction conditions; in addition 30 units of Calf Intestinal Phosphatase were added to the reaction to remove the phosphate from the 5' ends. The reaction was quenched by phenol/chloroform extraction, followed by ethanol-purification and resuspension in TE buffer. The SSNT oligonucleotide was phosphorylated at the 5' end with phage T4 polynucleotide kinase in the recommended reaction conditions.

Because of the low efficiency of phage T4 DNA ligase in ligating ssDNA to dsDNA, an oligonucleotide consisting of three or five adenine residues, depending on the number of thymine residues in SSNT, was added to the reaction. The poly(A) oligonucleotide did not ligate to the DNA fragments because of the absence of the 5' phosphate group. A typical ligation reaction was carried out in a volume of 100 μl as follows: 100 pmol of DNA fragments, 200 pmol of SSNT oligonucleotide, 500 pmol of poly(A), 50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 10 mM DTT, 100 mM NaCl, 1 mM ATP and 80 units of T4 DNA ligase. The reaction was incubated at 4°C for 16 hours. After heat-inactivation, 100 μl of 10 mM MgCl₂ was added to the reaction together with 30 units of NspI endonuclease (Boehringer Mannheim) followed by overnight incubation at 37°C. The 299 and 337 bp fragments, containing the ligated oligonucleotide, were gel-purified in 2% (w/v) agarose. The DNA was electroeluted by means of an Elutrap apparatus (Schleicher & Schuell, Keene NH), phenol/chloroform-extracted, ethanol-purified and resuspended in TE buffer. The two DNA fragments were then ligated in conditions similar to the previous ligation reaction. As before, an excess of poly(A) oligonucleotide (three or five A residues) was required for ligation of SSNT to the NspI site. Self-ligation of the 299 bp fragment could not be avoided but the difference in size between self-ligated and cross-ligated products allowed the 636 bp fragment with a single-stranded gap of poly(T) to be gel-purified in 2% agarose. The DNA was recovered by electroelution,
Because the poly(A) oligonucleotide, complementary to the poly(T) region of SSNT, was crucial for ligation efficiency, a slightly different strategy has been used to construct the DNA fragment with a 1 bp gap. As shown in Figure 7b, a 1 nt single-stranded gap was made by annealing three oligonucleotides with complementary sequences (SS1 T, LCOMP and RCOMP) and by ligating the construct to the 299 and 337 bp fragments. Again, for better yield, the ligation was carried out in two steps. The PCR amplification of the 720 bp DNA fragment and the phosphorylation of the 5' end of SS1T and LCOMP oligonucleotides were done as described above. The 720 DNA was digested overnight with NspI endonuclease in the recommended reaction conditions followed by heat-inactivation. Ligation of SS1T-LCOMP fragment to the NspI restriction site was done in a 400 µl reaction containing 400 pmol of NspI-digested DNA, 2 nmol each of SS1T and LCOMP oligonucleotides, 10 mM Tris-HCl (pH 7.9), 50 mM NaCl, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP, 25 µg/ml bovine serum albumin and 80 units of T4 ligase. The reaction was incubated overnight at 16°C and quenched by heat-inactivation. 50 units of XbaI endonuclease were added to the reaction followed by incubation at 37°C for four hours. The 299 bp fragment with the SS1T-LCOMP insert and the 337 bp DNA fragment were gel-purified in 2% agarose. Due to the co-migration of the two fragments, a single slice was excised from the gel and the DNA was recovered by electroelution. The DNA was phenol/chloroform-extracted, ethanol-precipitated and resuspended in TE buffer.

As shown in Figure 7b, annealing of the RCOMP oligonucleotide to the SS1T oligonucleotide inserted at the end of the 299 bp fragment produced a 5' overhang complementary to the XbaI site of the 337 bp fragment. However, because both sites carry a 5' phosphate group, self-ligation will significantly decrease the amount of cross-ligated product. This problem was overcome by designing the RCOMP oligonucleotide sequence in such a way that self-ligation of either fragments formed a StyI

---

**Figure 7(a)** (legend on page 56)
or a XbaI restriction site. Therefore, performing the ligation reaction in the presence of these two endonucleases allowed the formation of a single ligation product consisting of a 657 bp fragment with 1 nt single-stranded gap. Specifically, the ligation reaction was done in a volume of 200 \( \mu \)l as follows: 60 pmol of both 299-SS1T fragment and 337 bp fragment, 120 pmol of RCOMP oligonucleotide, 10 mM Tris-HCl (pH 7.9), 50 mM NaCl, 10 mM MgCl\(_2\), 10 mM DTT, 80 units of T4 DNA ligase, 10 units of XbaI, 10 units of Styl and 100 \( \mu \)g/ml bovine serum albumin. The reaction was incubated at 16°C for 15 hours and the 657 bp DNA construct was gel-purified and electroeluted as described above. The presence of a single-stranded gap in all the constructs was verified by denaturing PAGE of 5\( \mu \)l radiolabeled fragments.

The DNA fragment without any single-stranded gap was obtained by PCR amplification of the DNA fragment with a 5 nt single-stranded gap. The reaction was done under standard conditions using Deep Vent DNA polymerase and primers complementary to either ends. The 641 bp DNA fragment was gel-purified and electroeluted as described above.

Concentrations of DNA in all cases are reported in units of mol molecules and were determined by measuring absorbance at 260 nm. Enzymes were purchased from New England Biolabs unless otherwise specified.

**Scanning force microscopy**

The A-tract DNA samples were diluted to a concentration of 1 to 2 nM in buffer containing 4 mM Hepes (pH 7.4), 10 mM NaCl, 2 mM MgCl\(_2\). Immediately afterwards 10 to 20 \( \mu \)l of the DNA solution was deposited onto freshly cleaved ruby mica (Mica New York, NY). The sample was incubated for several minutes, rinsed with water and blown dry with nitrogen. The DNA samples with single-stranded gaps were deposited in the same manner, except that the deposition buffer contained MgCl\(_2\) at a concentration of 4 mM. All water was evaporated.
purified by distillation, followed by treatment in a Nano-pure water-purification apparatus (Barnstead, Dubuque Iowa). SFM images were obtained in air with a Nano-scope III microscope (Digital Instruments Inc., Santa Bar-bara, CA) operating in tapping mode. All operations were done at room temperature. Commercial diving board silicon tips (Nanosensor, Digital Instruments) were used. The microscope was equipped with a type E scanner (12 × 12 μm). The 512 × 512 pixels images were collected with a scan size of 1.5 μm at a scan rate varying between two and five scan lines per second.

Image processing

The SFM images were analyzed using ALEX, an image analysis toolbox written locally in the Matlab environment (MathWorks Inc., Natick MA). The image integer values of the Nanoscope file were converted to nanometers using the relation given in the Nanoscope III documentation. The images were flattened by subtracting from each scan line a least-squares fitted polynomial. No additional filter was applied to the images.

The DNA path was digitized as follows: the two DNA ends and several points along the DNA contour were selected with the mouse. The points were then interpolated with steps of one pixel. The DNA path was found by seeking, around each of the interpolated points, the highest intensity in a given window three to five pixels wide. To reduce the noise, the path trace was smoothed by polynomial fitting. The contour length was determined from the sum of the distances between consecutive points in the trace going from one end to the other. The end-to-end distance was defined as the distance between the first and the last points of each DNA path.

Monte Carlo simulation of worm-like chains with bends or regions of great flexibility

A simulated chain consisted of a series of n segments of length ℓ and infinitesimal thickness. The angle θ between two consecutive segments was chosen by a Monte Carlo method from normally distributed numbers with mean zero and variance of ℓ/|P|, where P is the persistence length. A-tract containing molecules with fixed bend angles were simulated by imposing an angle θ of 13.5° at locations corresponding to the beginning of the A-tract sequence in the real DNA molecules. The length of the individual sections was determined from the number of base-pairs using the experimental rise of 3.07 Å/bp. The persistence length was assumed to be 53 nm and the length ℓ of one segment was 3.07 Å. Molecules with flexible bends were simulated similarly, with the only difference being that the bend angle was introduced by adding 13.5° to the randomly chosen angle θ. SFM-like images were generated by passing a hypothetical tip with end radius of 5 nm over the simulated molecules (Keller & Franke, 1993). The height of the chains was chosen to be 2 nm.

Acknowledgments

We thank Professors Maxim Frank-Kamenetskii, Paul Hagerman, David Keller, Wilma Olson, Steve Smith and Victor Zhurkin for critical reading and comments on the manuscript. This work was supported by NSF grants MBC 9118482 and BIR 9318945 and NIH grant GM-32543. C.R. was supported by a Human Frontier Science Program (HFSP) long-term fellowship. C.W. was supported by NIH Training Grant GM 07759.

References


### Appendix

Equation (8) in the main text can be derived as follows. For convenience, let the unit vectors \( \vec{u}(\ell - \delta/2) = \vec{u}_- \) and \( \vec{u}(\ell + \delta/2) = \vec{u}_+ \) make an angle \( \beta \). Furthermore, let \( \vec{u}_- \) lie along the \( z \) axis of a standard Cartesian coordinate system. The unit vectors \( \vec{u}(s) \) and \( \vec{u}(s') \) can then be expressed in this coordinate system as:

\[
\vec{u}(s) = \sin \theta \cos \phi \vec{x} \\
\vec{u}(s') = \sin \theta' \cos \phi' \vec{x}
\]

\[
\sin \theta \sin \phi \vec{y} \\
\sin \theta' \sin \phi' \vec{y}
\]

\[
\cos \beta \vec{z}
\]

The average value of this quantity is therefore:

\[
\langle \vec{u}(s) \cdot \vec{u}(s') \rangle = \cos \theta \cos \theta' = \langle \vec{u}(s) \cdot \vec{u}(-\vec{u}_- \cdot \vec{u}(s')) \rangle
\]

Since the dot product is invariant under a change of coordinates, the unit vectors \( \vec{u}(s') \) and \( \vec{u}_- \) can be expressed in a coordinate system where \( \vec{u}_+ \) is now placed along the \( z \) axis. Recalling that \( \vec{u}_- \) and \( \vec{u}_+ \) form an angle \( \beta \) between them:

\[
\vec{u}(s') = \sin \theta' \cos \phi' \vec{x} \\
\vec{u}_- = \sin \beta \cos \phi \vec{x}
\]

\[
\sin \theta' \sin \phi' \vec{y} \\
\sin \beta \sin \phi \vec{y}
\]

\[
\cos \beta \vec{z}
\]
where the subscript $+$ indicates the coordinate system in which the $z$ axis coincides with the $\tilde{u}_+$ vector. Hence:

$$\langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle = \cos \beta \langle \tilde{u}(s) \cdot (\tilde{u}(\ell - \delta/2) \cdot \tilde{u}(s') \cdot \tilde{u}(\ell + \delta/2) \rangle \quad (A6)$$

Combining the results of equations (A3) and (A5), and using the definitions of $\tilde{u}_+$ and $\tilde{u}_-$ gives:

$$\langle \tilde{u}(s) \cdot \tilde{u}(s') \rangle = \cos \beta \langle \tilde{u}(s) \cdot \tilde{u}(\ell - \delta/2) \rangle \langle \tilde{u}(s') \cdot \tilde{u}(\ell + \delta/2) \rangle$$

thus, in the limit of $\delta \to 0$, equation (A6) yields equation (8) of the main text.