

# Characterization of the Geometry and Topology of DNA Pictured As a Discrete Collection of Atoms

Nicolas Clauvelin,<sup>\*,†</sup> Wilma K. Olson,<sup>†,‡</sup> and Irwin Tobias<sup>\*,‡</sup>

<sup>†</sup>BioMaPS Institute for Quantitative Biology and <sup>‡</sup>Department of Chemistry and Chemical Biology, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854, United States

**ABSTRACT:** The structural and physical properties of DNA are closely related to its geometry and topology. The classical mathematical treatment of DNA geometry and topology in terms of ideal smooth space curves was not designed to characterize the spatial arrangements of atoms found in high-resolution and simulated double-helical structures. We present here new and rigorous numerical methods for the rapid and accurate assessment of the geometry and topology of double-helical DNA structures in terms of the constituent atoms. These methods are well designed for large DNA data sets obtained in detailed numerical simulations or determined experimentally at high resolution. We illustrate the usefulness of our methodology by applying it to the analysis of three canonical double-helical DNA chains, a 65 bp minicircle obtained in recent molecular dynamics simulations, and a crystallographic array of protein-bound DNA duplexes. Although we focus on fully base-paired DNA structures, our methods can be extended to treat the geometry and topology of melted DNA structures as well as to characterize the folding of arbitrary molecules, such as RNA and cyclic peptides.

## 1. INTRODUCTION

The intertwining of the strands that bear the complementary bases of the double helix introduces topological properties in spatially constrained DNA molecules. Changes in the local base-pair structure, such as the melting that accompanies the transcription of the genetic message into RNA or the replication of the strands into daughter molecules, perturb the overall folding of covalently closed duplexes and looped fragments anchored at both ends to a protein scaffold. The interplay of local and global structure accordingly determines how DNA functions within the small, tightly packed confines of a cell.

Advances in structural and computational biology are now making it possible to decipher the effects of base-pair sequence, bound proteins, and small molecules on the local structure of DNA,<sup>1–5</sup> and state-of-the-art computations<sup>6–17</sup> are complementing these findings with direct examples of the influence of local sequence-, ligand-, and solvent-induced deformations of DNA helical structure on the spatial configurations and properties of increasingly longer chains. The rich details of molecular structure revealed in the numerical studies, ranging from the time-dependent positions of individual atoms collected in molecular dynamics calculations to the fluctuations of long molecular fragments extracted from Monte Carlo simulations, contrast sharply with the traditional mathematical treatment used to relate the local and global properties of double-helical DNA.

The traditional theory of DNA geometry and topology represents the double helix as a ribbon constructed from smooth curves describing an idealized structure.<sup>18</sup> The local structure is characterized in terms of the twisting of the edges of the ribbon about a curve through its center, called the centerline, and the global structure is described in terms of the folding of the centerline. The twisting of the ribbon is analogous to the winding of either strand of DNA about a smoothly deforming double-helical axis, and the pathway of the centerline is equivalent to the higher-order folding, or supercoiling, of the double-helical axis. A quantity Wr,

called the writhing number, or writhe, is a standard measure of the global configuration of a closed space curve. A second quantity Tw, referred to as the total twist, quantifies the twisting of a ribbon about its centerline. The sum of Wr and Tw, expressed as the number of turns of the ribbon, is referred to as the linking number Lk. Moreover, the value of Lk is an integer if the ends of the ribbon are brought together such that the edges form a continuous curve and the surface bounded by the edges of the ribbon is free of self-intersection.<sup>19–21</sup> Under conditions where the ribbon remains intact, changes in the global chiral folding, measured by the writhe, thus compensate for variations in the local twist. Such conditions hold in the case of DNA if the bases remain paired and the complementary strands have no covalent breaks.

Because DNA molecules are discrete objects, described in computations and experiments by collections of points rather than smooth curves, determination of the topological properties of the detailed structures now at hand requires careful adaptation of the traditional mathematical formalism. For example, given the nature of the dependence of Wr on the distances between the points along a smooth closed curve in the well-known Gauss integral expression,<sup>22,23</sup> approximation of the writhing number by a summation of corresponding terms fails if parts of the DNA come into close contact,<sup>24</sup> such as at the crossover points of braided or plectonemic configurations. Computation of the total twist of a discrete molecular pathway similarly requires an expression that yields an integer when combined with the writhing number of a closed structure. As shown recently,<sup>25</sup> the twist angle commonly used to describe the relative spatial arrangements of base-pair steps differs from the so-called twist of supercoiling, which is connected to the global pathway. As a consequence, one has to add to the writhing number the total twist of supercoiling to get the correct linking number.

**Received:** September 19, 2011

**Published:** January 20, 2012



Limitations on the computer resources needed to perform the demanding simulations of large DNA structures necessitate efficient as well as accurate determination of chain properties. As a first step in developing a set of methods to characterize the topology of the discrete structures generated in Monte Carlo and molecular dynamics simulations, here we present the mathematical framework needed to obtain exact, alternate expressions of  $Wr$  and  $Tw$  applicable to an arbitrary set of (smoothly connected or discrete) points and a general formalism that allows for the treatment of the wide variety of structural data generated in various numerical studies, such as the vertex positions and base-pair parameters used in different Monte Carlo studies,<sup>6,9,14,16</sup> the bead coordinates and orientations collected in Brownian dynamics simulations,<sup>7,12</sup> and the atomic coordinates reported in molecular dynamics studies.<sup>10,11,13,17</sup> Our methodology applies as well to the topological characterization of the growing number of DNA structures determined experimentally at high resolution and stored in the Nucleic Acid Database<sup>26</sup> and Protein Data Bank.<sup>27</sup> We limit attention, however, to double-helical structures and leave the more sophisticated methods needed to treat highly melted and other unusual DNA structures to a forthcoming publication.

Our goal is to bridge the gap between the traditional mathematical tools used to describe the geometry and topology of smooth curves and the discrete nature of DNA structures at the atomic scale. As we will show, the adaptation of the mathematical framework is not straightforward. One must define discrete quantities consistent with their smooth counterparts and, at the same time, derive a coarse-grained representation that captures well-known features of the double helix. To the best of our knowledge, we provide, for the first time, methods to evaluate the topological properties of DNA at the level of the constituent atoms. With such methods it becomes possible to observe whether structures determined experimentally or generated by numerical simulations resemble ideal models. One can also relate modifications in the atomic structure to global changes in the conformation and the folding of DNA.

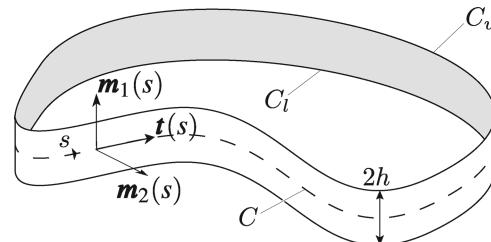
The cornerstone of our approach is the treatment of DNA as a discrete ribbon and the rigorous quantification of geometry and topology characteristic of this representation. The traditional perspective of DNA is easily extended to real three-dimensional structures within this framework and can be adapted to capture the changes in topology within individual strands, at localized sites, or in different chemical components of a covalently closed or spatially anchored duplex. We start by reviewing the classic ribbon model of DNA and the known definitions of ribbon topology. We next introduce our discrete ribbon representation of DNA and develop several new equations to describe the topology and geometry of a ribbon defined by an arbitrary set of points and the differences in topology between closely related structures. We then show how our approach can be applied to various types of DNA structures and present some numerical examples. We conclude with a discussion of the general applicability of the methodology to the treatment of the topology of arbitrary molecules, such as cyclic peptides and folded RNA.

## 2. PRELIMINARIES

As mentioned in Section 1, in our discussion of the topology of DNA, we shall show that the double-helical nature of DNA can always be made to be consistent with a ribbon representation.

Here we give a short overview of such a representation, first for the case of a smooth ribbon and then, in the next section, for the case of a discrete setting.

**2.1. Smooth Ribbon Geometry.** We consider a closed ribbon generated by starting with a curve having a so-called adapted frame—a centerline, corresponding to the central axis  $C$  of the ribbon, with a certain triad of mutually orthogonal unit vectors attached to it. The path of  $C$  is given by  $\mathbf{r}(s)$ , where  $s$  is the arc length, which is a parameter describing the position on the curve such that the derivative of  $\mathbf{r}(s)$  with respect to  $s$  gives a unit vector, and the frame attached to it is denoted  $\mathbf{m}(s) = \{\mathbf{m}_1(s), \mathbf{m}_2(s), \mathbf{t}(s)\}$  as illustrated in Figure 1. Note that the



**Figure 1.** Smooth ribbon geometry. The geometry of a smooth ribbon is described by its centerline  $C$ , its upper and lower edges,  $C_u$  and  $C_l$ , and a material frame,  $\mathbf{m}(s) = \{\mathbf{m}_1(s), \mathbf{m}_2(s), \mathbf{t}(s)\}$  attached to  $C$ , where  $\mathbf{t}(s)$  is the unit tangent of the centerline at arc length  $s$ . Both edges of the ribbon can be simply expressed in terms of the centerline,  $\mathbf{m}_1(s)$ , and the width of the ribbon.

frame is adapted to the centerline in the sense that  $\mathbf{m}_1(s)$  and  $\mathbf{m}_2(s)$  are not only normal to each other but also normal to the tangent  $\mathbf{t}(s) = d\mathbf{r}(s)/ds$  of the centerline, which is the third member of the triad. We also assume that the closed ribbon is nonintersecting, which means that the surface bound by the two edges of the ribbon is free of intersection with the two edge curves. As shown by White and Bauer,<sup>19</sup> this assumption is necessary to ensure that the topological properties of the ribbon, which will be detailed below, can be written in a simple form.

We specify, without any loss of generality, that the surface of the ribbon contains the unit vector  $\mathbf{m}_1(s)$ , which points toward  $C_u$ , the curve that is the upper edge of the ribbon, and that the unit vector  $\mathbf{m}_2(s)$  is normal to the surface of the ribbon as illustrated in Figure 1. Note that the upper edge of the ribbon is given by  $\mathbf{r}_u(s) = \mathbf{r}(s) + h\mathbf{m}_1(s)$ , where  $2h$  is the width of the ribbon, and the lower edge  $C_l$  is given by  $\mathbf{r}_l(s) = \mathbf{r}(s) - h\mathbf{m}_1(s)$ .

The kinematics of the ribbon is also characterized by the Darboux vector  $\Omega(s)$ , which quantifies the rate of spatial rotation of the material frame along the centerline. The magnitude of the component of  $\Omega(s)$  normal to the tangent of the centerline can be shown to be simply the curvature of  $C$ ,  $\kappa(s) = |d\mathbf{r}(s)/ds|$ , and its component,  $\tau(s)$ , along the tangent, corresponds to the twist density. The Darboux vector can then be written as

$$\Omega(s) = \kappa(s)\mathbf{b}(s) + \tau(s)\mathbf{t}(s) \quad (1)$$

where  $\mathbf{b}(s) = \mathbf{t}(s) \times \mathbf{t}'(s)/|\mathbf{t}(s) \times \mathbf{t}'(s)|$  is the binormal vector of the centerline. The twist density  $\tau(s)$  is given by

$$\tau(s) = \mathbf{m}'_1(s) \cdot \mathbf{m}_2(s) = \mathbf{t}(s) \cdot (\mathbf{m}_1(s) \times \mathbf{m}'_1(s)) \quad (2)$$

where the prime indicates differentiation with respect to arc length. In other words, the twist density of the ribbon is related to the twisting of the edges about the centerline. Finally we recall that the frame evolution along the centerline is given by

$\mathbf{m}'(s) = \Omega(s) \times \mathbf{m}(s)$ , where  $\mathbf{m}$  is any vector in the material frame  $\{\mathbf{m}_1(s), \mathbf{m}_2(s), \mathbf{t}(s)\}$ .

Following the work of Langer and Singer,<sup>28</sup> we can also introduce an adapted natural frame for the ribbon centerline (called the Bishop frame),<sup>29</sup> which will be useful within the discrete ribbon setting. This frame, denoted  $\{\mathbf{u}_B(s), \mathbf{v}_B(s), \mathbf{t}(s)\}$ , is different from the material frame in the sense that its propagation along the centerline is *twist-free*. The Darboux vector associated with this natural frame is given by  $\Omega_B(s) = \kappa(s)\mathbf{b}(s)$ , which implies that there is no tangential component and hence no twist. In other words this frame follows the centerline but the vectors  $\mathbf{u}_B(s)$  and  $\mathbf{v}_B(s)$  do not necessarily point toward the edge of the ribbon (in general, they do not). It then turns out that at any point on the centerline, the natural frame, and the material frame only differ by a rotation of angle  $\theta(s)$  about the centerline tangent  $\mathbf{t}(s)$ . We will refer to this angle as the integrated twist density and the rate of variation of this angle along the ribbon centerline as the twist density, that is,

$$\tau(s) = \frac{d}{ds}\theta(s) \quad (3)$$

For the closed ribbon, all of the functions of arc length defined in this section, except for the Bishop frame axes, are periodic: whatever value any such function has for a given value of  $s$ , it will have the same value when the arc length is set equal to  $s + L$ , where  $L$  is the length of the ribbon, for example,  $\mathbf{m}(0) = \mathbf{m}(L)$ .

**2.2. Smooth Ribbon Topology.** We next define the linking number, the writhing number, and the total twist and then describe the role these quantities play in a smooth closed ribbon.

The linking number, an integer, is a global quantity describing the entanglement of two closed space curves. It is, in fact, the algebraic sum of all of the signed intersections that one of the curves makes with a surface bounded by the other curve.<sup>18,20</sup> It is often referred to as a topological invariant<sup>21</sup> because during any deformation of the curves, the linking number remains constant as long as the curves do not intersect each other. For two nonintersecting closed curves  $C_1$  and  $C_2$  given by  $\mathbf{r}_1(s_1)$  and  $\mathbf{r}_2(s_2)$  the linking number, defined in the form of a Gauss integral,<sup>22,23</sup> is

$$\text{Lk}(C_2, C_1) = \frac{1}{4\pi} \oint \oint \frac{\mathbf{t}_2(s_2) \times \mathbf{t}_1(s_1) \cdot (\mathbf{r}_2(s_2) - \mathbf{r}_1(s_1))}{|\mathbf{r}_2(s_2) - \mathbf{r}_1(s_1)|^3} ds_1 ds_2 \quad (4)$$

where  $\mathbf{t}_1(s_1)$  and  $\mathbf{t}_2(s_2)$  are the unit tangents to  $C_1$  and  $C_2$ , respectively, and  $s_1$  and  $s_2$  are, respectively, the arc lengths of  $C_1$  and  $C_2$ .

The writhing number  $\text{Wr}(C)$  characterizes the global folding of a single closed curve  $C$ . It is a measure of the chiral distortion of that curve from planarity. As discussed by Fuller,<sup>30</sup> the writhe of a closed curve  $C$  can be defined as the number obtained by averaging, over all orientations of a plane  $P$  the directional writhe, which is the sum of the signed self-crossings that occur in the planar curves obtained from the perpendicular projection of  $C$  on  $P$ . The writhe can also be expressed as a Gauss integral,<sup>21,30,31</sup>

$$\text{Wr}(C) = \frac{1}{4\pi} \oint \oint \frac{\mathbf{t}(s') \times \mathbf{t}(s) \cdot (\mathbf{r}(s') - \mathbf{r}(s))}{|\mathbf{r}(s') - \mathbf{r}(s)|^3} ds ds' \quad (5)$$

where  $\mathbf{t}(s)$  the unit tangent to the curve  $C$ , and  $s$  and  $s'$  denote two positions on the curve.

Like the linking and writhing numbers, which are global quantities involving closed curves, the total twist of one curve,  $C_2$ , about the another,  $C_1$ , denoted  $\text{Tw}(C_2, C_1)$ , is also global, but the curves  $C_1$  and  $C_2$  need not be closed. The total twist can always be expressed as a single integral. In the case of the smooth ribbon having a width  $2h$ , we take  $C_1$  to be the centerline  $C$  given by  $\mathbf{r}(s)$  and the curve  $C_2$  to be the curve  $C_u$  traced out by the upper edge and given by  $\mathbf{r}_u(s) = \mathbf{r}(s) + h\mathbf{m}_1(s)$ . This twist (in units of number of turns) is simply the integral of  $\tau(s)$ , the twist density (eq 2), about the centerline, divided by  $2\pi$ , i.e.,

$$\begin{aligned} \text{Tw}(C_u, C) &= \frac{1}{2\pi} \oint \tau(s) ds \\ &= \frac{1}{2\pi} \oint \mathbf{t}(s) \cdot (\mathbf{m}_1(s) \times \mathbf{m}'_1(s)) ds \end{aligned} \quad (6)$$

Note that the value of the total twist does not depend on the value of  $h$  used to generate the curve  $C_u$  but only depends on the rate of rotation of  $C_u$  about  $C$ .

In the case of a closed smooth ribbon, if  $n$  is the integral number of turns of  $2\pi$  radians imposed on the original open flat ribbon before it is closed, then it is not difficult to show that the linking number of all pairs of ribbon curves of the form  $\mathbf{r}_1(s) = \mathbf{r}(s) + c_1\mathbf{m}_1(s)$  and  $\mathbf{r}_2(s) = \mathbf{r}(s) + c_2\mathbf{m}_1(s)$  is  $n$ . We have chosen the curve  $C_2$  to be the upper ribbon edge  $C_u$  ( $c_2 = h$ ) and curve  $C_1$  to be the centerline  $C$  ( $c_1 = 0$ ). The integer  $n$  is often referred to as the ribbon linking number. This linking number is unchanged if the ribbon is deformed into a new shape as long as, during the deformation, the ribbon is not broken, and there is no intersection of any of the curves on the ribbon.

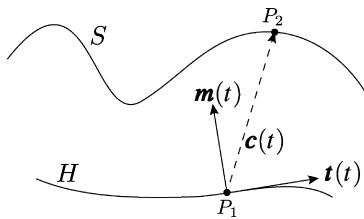
Not only is it well-known that the topology of a closed ribbon can be characterized by the integer  $n$ , the linking number of either one of the edges of the ribbon and its centerline, but also that the linking number can be expressed as the sum of the writhing number of the centerline and the total twist of an edge, here the upper edge, about the centerline<sup>18</sup>

$$\text{Lk}(C_u, C) = \text{Tw}(C_u, C) + \text{Wr}(C) \quad (7)$$

Similarly, when we consider how the value of the twist of the upper edge of the ribbon about the centerline changes during a ribbon deformation of the type mentioned above, in which the linking number is constant, it follows that the change in the value of the total twist,  $\delta\text{Tw}$ , is simply the negative of the change in the writhing number of the centerline,  $-\delta\text{Wr}$ .

**2.3. Ribbon Representation of Ideal DNA.** Several studies have addressed the ribbon representation of the DNA double helix. For example, White and Bauer<sup>32</sup> derived the centerline and the material frames of a DNA ribbon from a set of two smooth closed curves,  $H$  and  $S$ , that represent, respectively, the helical axis and one of the strands of an ideal covalently closed DNA double-helical molecule.

The vector  $\mathbf{c}(t)$  that connects corresponding points on the two closed curves, where  $t$  is a common parametrization of both curves, e.g.,  $P_1(t)$  located on curve  $H$  and  $P_2(t)$  located on curve  $S$ , as illustrated in Figure 2, is called the correspondence vector. When combined with the unit tangent of one of the curves, the correspondence vector defines the cross-section of the DNA ribbon, that is, the ribbon material frame. The material frame is defined by  $\mathbf{t}(t)$  the unit tangent of curve  $H$ ,  $\mathbf{m}_1(t)$  a unit vector perpendicular to  $\mathbf{t}(t)$  and lying in the plane of  $\mathbf{t}(t)$  and  $\mathbf{c}(t)$ , and



**Figure 2.** Correspondence vector and ribbon representation of points along two smooth curves. The correspondence vector  $c$  that connects the two curves  $H$  and  $S$  is used together with  $t$  the tangent of  $H$ , to define the material vector  $m$ . This method makes it possible to define a ribbon representation for the two curves in terms of  $H$  and  $S$ .

$\mathbf{m}_2(t)$  a unit vector obtained by the right-hand rule. The positive direction of  $\mathbf{m}_1(t)$  is taken along  $\mathbf{k}(t)$ , which is defined as

$$\mathbf{k}(t) = \mathbf{c}(t) - (\mathbf{c}(t) \cdot \mathbf{t}(t))\mathbf{t}(t)$$

The material frame attached to  $H$  is then given by

$$\{\mathbf{m}_1(t), \mathbf{m}_2(t), \mathbf{t}(t)\} = \left\{ \frac{\mathbf{k}(t)}{\|\mathbf{k}(t)\|}, \mathbf{t}(t) \times \frac{\mathbf{k}(t)}{\|\mathbf{k}(t)\|}, \mathbf{t}(t) \right\}$$

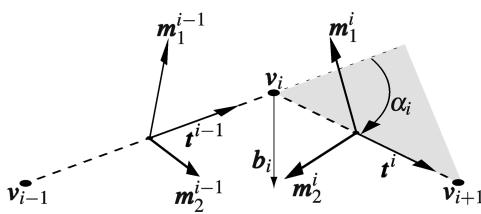
This frame represents a material frame in the sense that the director vectors,  $\mathbf{m}_1$  and  $\mathbf{m}_2$ , follow the twist of  $S$  around  $H$ .

The total twist of  $S$  about  $H$  can then be evaluated by the use of this material frame (eq 6). However, as discussed by White and Bauer,<sup>32</sup> even though the material frame is based on a particular definition of the correspondence vector, the total twist does not depend on the choice of mapping between curves  $H$  and  $S$ .

### 3. DISCRETE MODEL FOR RIBBON GEOMETRY AND TOPOLOGY

We begin this section with a mathematical description of the centerline with its attached material frame for a closed discrete ribbon and then move on to a discussion of the geometry and the topology of such a ribbon.

**3.1. Discrete Setting.** The centerline of the discrete ribbon consists of a set of  $N$  vertices  $\mathbf{v}_i$  with  $i = 1, \dots, N$ . Adjacent vertices are connected by line segments  $\mathbf{e}^i = \mathbf{v}_{i+1} - \mathbf{v}_i$ , each having as its tangent,  $\mathbf{t}^i = \mathbf{e}^i / \|\mathbf{e}^i\|$ , as illustrated in Figure 3. (Here



**Figure 3.** Discrete ribbon geometry. Within our discrete setting the ribbon centerline is given by a set of  $N$  vertices  $\{\mathbf{v}_i\}$ . A material frame  $\underline{\mathbf{m}}^i = \{\mathbf{m}_1^i, \mathbf{m}_2^i, \mathbf{t}^i\}$  is attached at the midpoint of each line segment  $\mathbf{e}^i = \mathbf{v}_{i+1} - \mathbf{v}_i$  along the centerline. The vector  $\mathbf{b}_i = \mathbf{t}^{i-1} \times \mathbf{t}^i / \|\mathbf{t}^{i-1} \times \mathbf{t}^i\|$  is the binormal and  $\alpha_i$  is the turning angle between the tangents  $\mathbf{t}^{i-1}$  and  $\mathbf{t}^i$ .

and below, symbols bearing upper indices denote segment-based quantities, while those with lower indices indicate properties of vertices.) There is a binormal vector  $\mathbf{b}_i$  associated with the  $i$ -th vertex, a unit vector in the direction of  $\mathbf{t}^{i-1} \times \mathbf{t}^i$ . We denote the

angle between the incoming and outgoing tangents as  $\alpha_i$  (see Figure 3). The adapted discrete material frame, denoted  $\underline{\mathbf{m}}^i = \{\mathbf{m}_1^i, \mathbf{m}_2^i, \mathbf{t}^i\}$ , is attached to the midpoint of each segment of the centerline. A discrete ribbon configuration is fully specified by the values  $\{\mathbf{v}_i\}$  and  $\{\underline{\mathbf{m}}^i\}$  and henceforth denoted  $\{\mathbf{v}_i; \underline{\mathbf{m}}^i\}$ .

Given the segmented centerline with its set of material frames, an approach along the lines of the method used to define the smooth ribbon can be used to determine the positions of the vertices that make up the upper and lower edges of the discrete ribbon. Here these vertices are respectively defined by  $(\mathbf{v}_i + \mathbf{v}_{i+1})/2 + h\underline{\mathbf{m}}_1^i$  and  $(\mathbf{v}_i + \mathbf{v}_{i+1})/2 - h\underline{\mathbf{m}}_1^i$ , where  $h$  is half the width of the ribbon.

As with the smooth ribbon, the periodic nature of the closed discrete ribbon gives rise to the fact that for any segment or vertex property, adding  $N$  to the position label does not change the position or property.

**3.2. Geometry.** We now present methods to evaluate the curvature and the twist density of the discrete ribbon.

**3.2.1. Discrete Curvature.** Although the curvature of the segmented centerline on the discrete ribbon is zero everywhere except at the vertices, where it is infinite, it is possible to derive a formula for the integral of the curvature and the average curvature associated with each vertex. Further insight into the nature of this type of centerline curvature is given in Appendices A.1 and A.2.

We define the length associated with each vertex, denoted  $l_i$ , as<sup>33,34</sup>

$$l_i = \frac{|\mathbf{e}^{i-1}| + |\mathbf{e}^i|}{2} \quad (8)$$

This length can be regarded as the integration domain attached to each vertex<sup>33</sup> and is used to define average vertex-based quantities.

As shown in Appendix A.2, the integrated curvature across a vertex is equal to the turning angle at this vertex. The average curvature associated with the  $i$ -th vertex, denoted as  $\bar{\kappa}_i$ , is then written as the turning angle at this vertex divided by the associated vertex length, that is

$$\bar{\kappa}_i = \frac{\alpha_i}{l_i} \quad (9)$$

which, for simplicity, we will refer to as the curvature in the examples below.

**3.2.2. Discrete Twist Density.** As discussed in Section 2, we can take advantage of the elegant work of Langer and Singer<sup>28</sup> to compute the twist density associated with a given vertex. Our goal is to provide a discrete analog of that expression (eq 3) along the lines of the work of Bergou et al.<sup>33</sup>

Considering the  $i$ -th vertex  $\mathbf{v}_i$  and the two adjacent material frames  $\underline{\mathbf{m}}^{i-1}$  and  $\underline{\mathbf{m}}^i$ , we introduce the rotation operator  $\underline{\mathbf{R}}_i(\mathbf{b}_i, \alpha_i)$  defined by

$$\underline{\mathbf{R}}_i(\mathbf{b}_i, \alpha_i) \cdot \mathbf{t}^{i-1} = \mathbf{t}^i \quad \text{and} \quad \underline{\mathbf{R}}_i(\mathbf{b}_i, \alpha_i) \cdot \mathbf{b}_i = \mathbf{b}_i \quad (10)$$

This operator is the rotation about the binormal at the  $i$ -th vertex by an angle  $\alpha_i$ <sup>35</sup> which is the turning angle between the two tangents (see Figure 3). The material frame  $\underline{\mathbf{m}}^{i-1}$  defined on the  $(i-1)$ -th segment can be transported onto the next  $i$ -th segment by applying this operator. The transported frame  $\underline{\mathbf{m}}^{i*}$  is then given by  $\underline{\mathbf{m}}^{i*} = \underline{\mathbf{R}}_i(\mathbf{b}_i, \alpha_i) \cdot \underline{\mathbf{m}}^{i-1}$ . This transport step is equivalent to the *twist-free* propagation of the adapted natural frame in the smooth setting. The angle  $\theta_i$  between the two frames  $\underline{\mathbf{m}}^{i*}$  and  $\underline{\mathbf{m}}^i$  can thus be computed with the use of the

following equations:

$$\begin{cases} \cos \theta_i = \mathbf{m}_1^i \star \cdot \mathbf{m}_1^i \\ \sin \theta_i = \mathbf{t}^i \cdot (\mathbf{m}_1^i \star \times \mathbf{m}_1^i) \end{cases} \quad (11)$$

where  $\theta_i$  is always taken in the range  $[-\pi; \pi]$ . The angle  $\theta_i$  corresponds to the integrated twist density along the vertex domain. The average discrete twist density at the vertex  $\mathbf{v}_i$  can accordingly be expressed as<sup>33</sup>

$$\bar{\tau}_i = \frac{\theta_i}{l_i} \quad (12)$$

which we will refer to as the twist density in the examples below. This equation is the discrete counterpart of the definition of the twist density given above for the smooth ribbon (eq 3). Since  $\mathbf{b}_i$  is normal to both  $\mathbf{t}^{i-1}$  and  $\mathbf{t}^i$ , the sum of the angles between  $\mathbf{m}_1^{i-1}$  and  $\mathbf{b}_i$  and between  $\mathbf{b}_i$  and  $\mathbf{m}_1^i$  leads to the same value for the integrated twist density  $\theta_i$ .

Our definition of the twist density as a vertex quantity differs from the approach taken by Britton et al.,<sup>25</sup> in which the twist density is associated with a segment. In the cited reference the vertices lay at the centers of the DNA base pairs. Each  $\mathbf{m}_1^i$  vector, the direction of which depended on the orientation of the associated base-pair plane about its normal, emanated from a vertex and was perpendicular to a unit vector pointing in the direction of the sum of the incoming and outgoing tangents. The twist, in this case ascribed to a pair of adjacent base pairs, is the sum of the three angles specified in Appendix B.

**3.3. Topology.** Now that we have introduced our discrete ribbon setting and its geometry, we present several methods to compute the topology of a segmented ribbon. We first present the definitions of the linking number, the writhing number, and the total twist of the discrete ribbon and then derive some methods to quantify changes in ribbon topology.

**3.3.1. Linking Number: Direct Evaluation.** We start with a method for the direct integration of the Gauss integral for the linking number (eq 4) for two segmented curves.

Our method is an extension of the method discussed in the papers of Swigon et al.<sup>36</sup> and Britton et al.<sup>25</sup> for the computation of the writhing number of a segmented curve. The very same method used in those papers can be used to write the linking number of two discrete curves as a sum of dihedral angles computed between all pairs of segments, one from each curve. For our discrete ribbon, the two curves are the centerline and one of the edge curves (we choose the upper edge curve) that can be constructed from the vertices  $\{\mathbf{v}_i\}$  and the material frames  $\{\mathbf{m}_1^i\}$ .

The two segments  $(\mathbf{v}_i, \mathbf{v}_{i+1})$  on the first curve and  $(\mathbf{v}_j, \mathbf{v}_{j+1})$  on the second curve define four different dihedral angles. Each angle corresponds to the angle between the normals of the two planes formed by the segments and a joining vector. The joining vector between two points (one on each segment), with relative displacements  $s^i$  and  $s^j$ , both in the range  $[0;1]$ , is defined as

$$\mathbf{r}_{ij}(s^i, s^j) = \mathbf{v}_i + s^i \mathbf{e}^i - \mathbf{v}_j - s^j \mathbf{e}^j \quad (13)$$

The two planes, one formed by the segment  $(\mathbf{v}_i, \mathbf{v}_{i+1})$  and  $\mathbf{r}_{ij}(s^i, s^j)$  and a second one formed by  $(\mathbf{v}_j, \mathbf{v}_{j+1})$  and  $\mathbf{r}_{ij}(s^i, s^j)$ , define the

dihedral angle  $\xi^{ij}(s^i, s^j)$  with the following cosine and sine:

$$\begin{cases} \cos \xi^{ij}(s^i, s^j) = \frac{\mathbf{r}_{ij}(s^i, s^j) \times \mathbf{t}^j \cdot \mathbf{t}^i \times \mathbf{r}_{ij}(s^i, s^j)}{|\mathbf{r}_{ij}(s^i, s^j) \times \mathbf{t}^j| |\mathbf{t}^i \times \mathbf{r}_{ij}(s^i, s^j)|} \\ \sin \xi^{ij}(s^i, s^j) = - \frac{\mathbf{r}_{ij}(s^i, s^j)}{|\mathbf{r}_{ij}(s^i, s^j)|} \cdot \left( \frac{\mathbf{r}_{ij}(s^i, s^j) \times \mathbf{t}^j}{|\mathbf{r}_{ij}(s^i, s^j) \times \mathbf{t}^j|} \right. \\ \left. \times \frac{\mathbf{t}^i \times \mathbf{r}_{ij}(s^i, s^j)}{|\mathbf{t}^i \times \mathbf{r}_{ij}(s^i, s^j)|} \right) \end{cases} \quad (14)$$

where  $\mathbf{t}^i$  and  $\mathbf{t}^j$  are the respective tangents of the first and the second segments.

The expression obtained for the writhing number of a segmented curve by Britton et al.<sup>25</sup> then implies that the linking number of a discrete ribbon configuration  $\text{Lk}\{\mathbf{v}_i; \mathbf{m}_1^i\}$  is

$$\text{Lk}(\{\mathbf{v}_i; \mathbf{m}_1^i\}) = \frac{1}{4\pi} \sum_{(p,q)} \text{lk}^{pq} \quad (15)$$

where the  $(p, q)$  pairs are such that segment  $p$  is part of the centerline and segment  $q$  is part of the traced-out material curve. The contribution of each pair is of the form:

$$\text{lk}^{pq} = \xi^{pq}(1, 0) + \xi^{pq}(0, 1) - \xi^{pq}(1, 1) - \xi^{pq}(0, 0) \quad (16)$$

Although the evaluation of the linking number of two segmented curves with this expression is costly from a computational point of view, it has the advantage of being direct and exact (up to the numerical precision of the computation).

**3.3.2. Writhing Number.** As mentioned above, the writhing number of a discrete ribbon can be computed using the methods of Swigon et al.<sup>36</sup> and Britton et al.<sup>25</sup> Klenin and Langowski<sup>37</sup> have also addressed the writhing number of a segmented DNA curve, along the same lines as Levitt,<sup>24</sup> as a sum of four angles. The latter angles are related to the solid angles of viewing directions along which all pairs of line segments intersect and thus are different from the dihedral angles employed here.

The writhing number of a closed discrete curve, in our case the centerline of the discrete ribbon, can be written as

$$\text{Wr}(\{\mathbf{v}_i\}) = \frac{1}{2\pi} \sum_{(p,q), q > p} w^{pq} \quad (17)$$

where the sum is taken over all pairs of centerline segments such that double counting is avoided. The contribution  $w^{pq}$  to the sum for the pair of segments defined by  $p$  and  $q$  is

$$w^{pq} = \xi^{pq}(1, 0) + \xi^{pq}(0, 1) - \xi^{pq}(1, 1) - \xi^{pq}(0, 0) \quad (18)$$

where  $\xi^{pq}(s^p, s^q)$  is obtained using the dihedral angle definition described above (eq 14).

**3.3.3. Total Twist.** The total twist in the smooth setting is an integral of the twist density (eq 2). In the discrete setting the total twist can be expressed simply as the sum,

$$\text{Tw}(\{\mathbf{v}_i; \mathbf{m}_1^i\}) = \frac{1}{2\pi} \sum_i \theta_i \quad (19)$$

where the angle  $\theta_i$  corresponds to the integrated twist density (eq 11). We recall that by the total twist of a ribbon we mean the total twist of one of the ribbon edges about the ribbon centerline.

**3.3.4. Linking Number Decomposition.** Given the above expressions for the writhing number (eq 17) and the total twist (eq 19) of a closed ribbon within our discrete setting, we can

also obtain the linking number of the ribbon, that is the linking number of one of the edges and the ribbon centerline, by adding the writhing number to the total twist (eq 7):

$$\text{Lk}(\{\mathbf{v}_i; \underline{\mathbf{m}}^i\}) = \text{Tw}(\{\mathbf{v}_i; \underline{\mathbf{m}}^i\}) + \text{Wr}(\{\mathbf{v}_i\}) \quad (20)$$

We emphasize the fact that the methods discussed here to determine the writhing number and the total twist within our discrete ribbon setting do not lead to an approximation of the linking number: the sum of the writhing number and the total twist always gives an integer<sup>28,38</sup> (up to the precision of the computation).

**3.4. Computing Changes in the Topology.** Numerical simulations of DNA configurations commonly entail incremental changes in the system (such as changes in the position and the orientation of the system components). For a new configuration, obtained by the application of incremental changes to an original configuration, it is always possible to evaluate the topology for each configuration on its own. However, it is often preferable to compute directly the variation in the topology induced by those incremental changes. In the case of closed double-helical DNA, it is possible to take advantage of the conservation of the linking number to obtain efficient computational methods to track the changes in the topology of the molecule.

We consider a discrete closed ribbon configuration  $\{\mathbf{v}'_i; \underline{\mathbf{m}}'^i\}$  obtained from an original configuration  $\{\mathbf{v}_i; \underline{\mathbf{m}}^i\}$  by applying some incremental changes, meaning that we have in symbolic form

$$\{\mathbf{v}'_i; \underline{\mathbf{m}}'^i\} = \{\mathbf{v}_i; \underline{\mathbf{m}}^i\} + \{\delta\mathbf{v}_i; \delta\underline{\mathbf{m}}^i\} \quad (21)$$

where  $\delta\mathbf{v}_i$  and  $\delta\underline{\mathbf{m}}^i$  denote, respectively, the changes in the vertex positions and in the material frame orientations. If the new configuration is obtained in a way that ensures the conservation of the linking number, meaning that  $\text{Lk}(\{\mathbf{v}_i; \underline{\mathbf{m}}^i\}) = \text{Lk}(\{\mathbf{v}'_i; \underline{\mathbf{m}}'^i\})$ , then

$$\delta\text{Wr} = -\delta\text{Tw} \quad (22)$$

where  $\delta\text{Wr} = \text{Wr}(\{\mathbf{v}'_i\}) - \text{Wr}(\{\mathbf{v}_i\})$  and  $\delta\text{Tw} = \text{Tw}(\{\mathbf{v}'_i; \underline{\mathbf{m}}'^i\}) - \text{Tw}(\{\mathbf{v}_i; \underline{\mathbf{m}}^i\})$ . Notice that this equality is strictly equivalent to the one derived in the smooth case.

Several researchers have already noticed the importance of this conservation equation either in theoretical models<sup>18</sup> or in computational methods.<sup>37,39,40</sup> We present here some novel methods to compute efficiently the changes in topology.

**3.4.1. Writhing Number Gradient.** As reported by Fuller,<sup>18</sup> it is possible to write the writh difference between two smooth closed curves  $C_1$  and  $C_2$  as a single integral

$$\begin{aligned} \Delta\text{Wr} &= \text{Wr}(C_2) - \text{Wr}(C_1) \\ &= \frac{1}{2\pi} \oint \frac{\mathbf{t}_1(t) \times \mathbf{t}_2(t)}{1 + \mathbf{t}_1(t) \cdot \mathbf{t}_2(t)} \frac{d}{dt} (\mathbf{t}_1(t) + \mathbf{t}_2(t)) dt \end{aligned} \quad (23)$$

where  $\mathbf{t}_1(t)$  and  $\mathbf{t}_2(t)$  are the unit tangents of  $C_1$  and  $C_2$  and  $t$  is a parametrization of both curves. This theorem implies several assumptions about the nature of  $C_1$  and  $C_2$  and the reader is referred to the work of Aldinger et al.<sup>41</sup> and Neukirch and Starostin<sup>42</sup> for detailed discussions.

As shown in Appendix A.3, the Fuller formula can be linearized for infinitesimal changes in the vertex positions. We consider the discrete ribbon centerline  $C$  defined by the set of vertices  $\{\mathbf{v}_i\}$  and for each vertex we introduce a translation noted  $\delta\mathbf{v}_i$ . A new ribbon centerline  $C'$  is then obtained when we consider the vertices  $\{\mathbf{v}'_i\} = \{\mathbf{v}_i + \delta\mathbf{v}_i\}$ . If these changes are small, we can use the formula

obtained in Appendix A.3 to compute the change in the writhing number of the ribbon, that is, the writhing number of the centerline, as

$$\delta\text{Wr} = -\frac{1}{2\pi} \sum_i \tan\left(\frac{\alpha_i}{2}\right) \mathbf{b}_i \cdot \left( \frac{\delta\mathbf{v}_{i+1} - \delta\mathbf{v}_i}{|\mathbf{e}^i|} + \frac{\delta\mathbf{v}_i - \delta\mathbf{v}_{i-1}}{|\mathbf{e}^{i-1}|} \right) \quad (24)$$

where  $\mathbf{b}_i$  is the binormal vector at the  $i$ -th vertex. The linearization underlying this result is valid as long as the changes in the vertex positions are small compared to the segment lengths, that is  $|\delta\mathbf{v}_i| \ll |\mathbf{e}^i|$  and  $|\delta\mathbf{v}_i| \ll |\mathbf{e}^{i-1}|$ .

A related formula is that for the gradient of the writhing number with respect to the vertex positions, which is written as

$$\begin{aligned} \frac{d}{d\mathbf{v}_i} \text{Wr}(\{\mathbf{v}_i\}) &= -\frac{1}{2\pi} \left[ \frac{\mathbf{b}_{i-1}}{|\mathbf{e}^{i-1}|} \tan \frac{\alpha_{i-1}}{2} + \left( \frac{\mathbf{b}_i}{|\mathbf{e}^i|} - \frac{\mathbf{b}_{i-1}}{|\mathbf{e}^{i-1}|} \right) \tan \frac{\alpha_i}{2} \right. \\ &\quad \left. - \frac{\mathbf{b}_{i+1}}{|\mathbf{e}^i|} \tan \frac{\alpha_{i+1}}{2} \right] \end{aligned} \quad (25)$$

As any gradient, this vector points in the direction in which a given infinitesimal shift of the  $i$ -th vertex will result in the largest possible increase in the writhe.

The preceding two expressions make it possible to obtain the difference in writhe, and hence the change in the total twist, for a discrete ribbon undergoing small configurational changes. These results are especially well suited to molecular dynamics simulations, in which the positions of atoms are subjected to small displacements at each time step.

**3.4.2. Writh Increment.** It is also possible to compute the increment in the writhe of a discrete ribbon for finite changes. To do so, we first have to introduce the self-linking number of the ribbon centerline: the self-linking number of a curve is defined here as the linking number of the curve itself with the curve traced out by the binormal vector. In our discrete setting, we have the following decomposition for the self-linking number

$$\text{SLk}(\{\mathbf{v}_i\}) = \text{Wr}(\{\mathbf{v}_i\}) + \text{Tw}(\{\mathbf{v}_i; \underline{\mathbf{m}}_{b,i}\}) \quad (26)$$

where the  $\underline{\mathbf{m}}_{b,i}$  are orthogonal frames located at each vertex and built using the binormal vector, that is

$$\begin{aligned} \underline{\mathbf{m}}_{b,i} &= \{\mathbf{b}_i, (\mathbf{t}^{i-1} + \mathbf{t}^i)/|\mathbf{t}^{i-1} + \mathbf{t}^i| \times \mathbf{b}_i, \\ &\quad (\mathbf{t}^{i-1} + \mathbf{t}^i)/|\mathbf{t}^{i-1} + \mathbf{t}^i|\} \end{aligned} \quad (27)$$

We emphasize the point that the self-linking number can be evaluated with the method introduced above for the direct evaluation of the linking number of two segmented curves (eq 15).

We also note that, contrary to the linking number of two curves, the self-linking number of a curve is not necessarily conserved under an arbitrary transformation.<sup>31</sup> However, like the linking number, the self-linking number is an integer, and hence its variation can only be an integer. Within our notation we can write

$$\begin{aligned} \text{Wr}(\{\mathbf{v}'_i\}) - \text{Wr}(\{\mathbf{v}_i\}) &= \text{SLk}(\{\mathbf{v}'_i\}) - \text{SLk}(\{\mathbf{v}_i\}) \\ &\quad - (\text{Tw}(\{\mathbf{v}'_i; \underline{\mathbf{m}}'_{b,i}\}) - \text{Tw}(\{\mathbf{v}_i; \underline{\mathbf{m}}_{b,i}\})) \end{aligned} \quad (28)$$

It is straightforward to show that the computation of the increment in the total twist in this case is equivalent to the determination of the change in the angles between successive binormal vectors. We then rewrite the writhe increment as

$$\text{Wr}(\{\mathbf{v}'_i\}) - \text{Wr}(\{\mathbf{v}_i\}) = -\frac{1}{2\pi} \sum_i (\phi'^i - \phi^i) + \Delta\text{SLk} \quad (29)$$

where  $\phi^i$  is the angle between the vectors  $\mathbf{b}_i$  and  $\mathbf{b}_{i+1}$  on the centerline of the original ribbon,  $\phi'^i$  is the corresponding angle on

the centerline of the modified ribbon centerline, and  $\Delta S_{Lk}$  is the variation in the self-linking number. As shown in Appendix B, the angle  $\phi^i$  is defined by the following cosine and sine

$$\begin{cases} \cos \phi^i = \mathbf{b}_i \cdot \mathbf{b}_{i+1} \\ \sin \phi^i = (\mathbf{b}_i \times \mathbf{b}_{i+1}) \cdot \mathbf{t}^i \end{cases} \quad (30)$$

This last result has the advantage of making it possible to compute the difference in the writhe for finite changes in the vertex positions when the new material frames are not known. Although one can argue that this formula introduces the costly evaluation of the self-linking number of the ribbon centerline, one can expect to conserve the self-linking number for small finite changes. Moreover, it is also possible to assume that small changes in the vertex positions induce small changes in the writhe. Then, if the writhe increment, computed with the assumption that the self-linking number is conserved, does not lie between  $-1$  and  $1$ , it simply means that the self-linking number has changed and that  $1$  or  $-1$  should be respectively subtracted from the writhe increment.

**3.4.3. Twist Increment.** As a direct consequence of the linking number conservation, the total twist increment is known as soon as the writhe increment is known. However, if the material frames are known for the modified configuration it is usually faster to recompute the total twist and then the twist increment (in which case the writhe increment is known since the linking number is conserved).

This is particularly true in Monte Carlo simulations of closed DNA in which a specific type of move is used to generate a new configuration: the move involves the random choice of two vertices that define an axis about which the shortest part of the closed DNA between the two vertices is rotated by a random angle.<sup>5,43</sup> Such a transformation of the discrete ribbon gives rise to a change in the twist density only for the two vertices that define the axis of rotation. The computation of these two new values is sufficient to obtain the increment in the total twist and hence the writhe increment.

#### 4. RIBBON REPRESENTATIONS AND DNA SIMULATIONS

The increase in computing resources and biophysical data worldwide has led to an upsurge of interest in numerical simulations of the structures of DNA molecules. The methods developed here for the analysis of DNA topology were designed to be compatible with the computational approaches most widely used to model DNA, namely molecular dynamics, Monte Carlo, and Brownian dynamics methods.

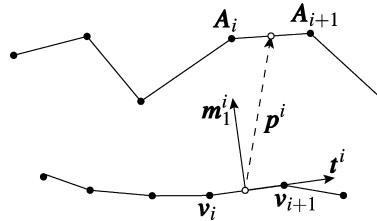
Molecular dynamics simulations of DNA model the physical movements of the constituent atoms over time. One outcome of such simulations is a detailed description of the configuration of the molecule, meaning the positions and the velocities of all atoms, at each time step. The methods we present here are designed to deal with such data and with the more coarse-grained representations of DNA structure used in other molecular simulations. We describe a way to construct a ribbon representation from a set of atomic coordinates and then we use our approach to characterize the geometry and topology of several structures, including: (i) three ideal double-helical models deduced from X-ray diffraction measurements;<sup>44,45</sup> (ii) a 65 bp DNA minicircle captured in recent molecular dynamics simulations;<sup>17</sup> and (iii) an array of closely packed protein-bound double-helical

fragments found in the high-resolution crystal structure of DNA complexed to the *Borrelia burgdorferi* Hbb protein.<sup>46</sup>

#### 4.1. Ribbon Representation and Atomic Coordinates.

We start with a simple method to build a ribbon representation from atomic coordinates. Although this example deals with discrete data (atomic coordinates), the approach is very similar to that developed by White and Bauer<sup>32</sup> for smooth curves and discussed in the beginning of this paper.

We consider two sets of atomic coordinates: one denoted  $\{\mathbf{v}_i\}$ , which represents the centerline of the ribbon, and a second one denoted  $\{\mathbf{A}_i\}$ , which is used to derive the material frames (see Figure 4). Following the method introduced in



**Figure 4.** Ribbon representation and atomic coordinates. A ribbon representation of a molecular structure can be obtained by projection of the coordinates of one set of atoms  $\{\mathbf{A}_i\}$  on the segmented pathway defined by a second set  $\{\mathbf{v}_i\}$ . The latter atoms serve as vertices  $\{\mathbf{v}_i\}$  on the centerline of the ribbon, and the vector  $\mathbf{p}^i$  that connects the midpoints of the segments  $(\mathbf{v}_i, \mathbf{v}_{i+1})$  and  $(\mathbf{A}_i, \mathbf{A}_{i+1})$  determines the plane containing the tangent  $\mathbf{t}^i$  and the first director vector  $\mathbf{m}_1^i$ . The direction of  $\mathbf{m}_1^i$  is taken along the normal component of  $\mathbf{p}^i$  with respect to the tangent  $\mathbf{t}^i$ .

Section 2.3, we define a joining vector  $\mathbf{p}^i$  between the midpoints of corresponding segments in the two sets of atoms,  $(\mathbf{v}_i, \mathbf{v}_{i+1})$  and  $(\mathbf{A}_i, \mathbf{A}_{i+1})$ . We take the tangent  $\mathbf{t}^i$  to lie along the segment containing  $(\mathbf{v}_i, \mathbf{v}_{i+1})$  and define the director vector  $\mathbf{m}_1^i$  as the normalized normal component of  $\mathbf{p}^i$  with respect to the tangent  $\mathbf{t}^i$ , that is,

$$\mathbf{m}_1^i = \frac{\mathbf{p}^i - \mathbf{t}^i(\mathbf{p}^i \cdot \mathbf{t}^i)}{|\mathbf{p}^i - \mathbf{t}^i(\mathbf{p}^i \cdot \mathbf{t}^i)|} \quad (31)$$

The second director vector of the material frame is then obtained by the right-hand rule,  $\mathbf{m}_2^i = \mathbf{t}^i \times \mathbf{m}_1^i$ .

The geometry and the topology of the ribbon defined by the selected atoms can be characterized with the expressions described above.

Obviously, in the case of a closed double-stranded DNA molecule there are many different choices of atoms for constructing a ribbon. We will discuss different possible choices below, in the context of the geometry and the topology of a DNA minicircle.

**4.2. Application to the Geometry and the Topology of Ideal DNA Helices.** We first consider 20 bp fragments of three ideal DNA double helices—A, B, and C DNA, with, respectively, 11, 10, and 9 base pairs per helical turn—that are representative of the ways in which the right-handed double-helical structure responds to changes in its chemical environment.<sup>47–49</sup> We choose these examples to give the reader insights into familiar forms of DNA and to provide useful standards for the analysis of other structures. The atomic coordinates for these structures have been obtained using the 3DNA software,<sup>50</sup> each with the sequence (ACTGA)<sub>4</sub> but with the different conformational repeating patterns found in the canonical X-ray diffraction models.<sup>44,45</sup> As explained below, we

compute the geometry and the topology of these structures by defining a discrete ribbon for each DNA strand and for the double helix.

**4.2.1. Ribbons Definitions.** The ribbon for a single strand is constructed here from the phosphorus P atoms and the glycosyl nitrogen N atoms that serve as anchor points for the bases on the sugar–phosphate backbone (N9 for adenine and guanine and N1 for thymine and cytosine). The P atoms, taken in sequential order along the 5'-3' direction of the strand, specify the centerline, and the N9/N1 atoms of the corresponding nucleotides define the material frames. Once the discrete ribbon for the strand is obtained, we can compute the average curvature (eq 9) and the average twist density (eq 12) at the vertices. Since we consider open DNA fragments, we cannot compute a linking or a writhing number; however, we can compute the total twist of the ribbon and the inclination of the ribbon with respect to the global double-helix axis of revolution. The ribbon inclination angle  $\psi$  is defined as the angle between the first material frame director,  $\mathbf{m}_1$ , and the ideal DNA double-helical axis taken as  $Z = (0,0,1)$  (for the second strand the direction of the double-helix axis is reversed). Finally, we note that since the ideal structures are composed of regularly repeating nucleotide units, the values of the geometric parameters along the ribbon are constant, and the choice of strand does not influence the results.

We can also address the geometry of the DNA fragments at the double-helix level by reconstructing a ribbon based on the atoms in the two strands. The centerline of this ribbon is defined by the average positions of two corresponding phosphorus P atoms, i.e., the average positions of the P atoms belonging to the nucleotides that form a base pair. Either set of phosphorus atoms (since we average the strands, choosing one strand versus the other does not change the results) can be used with the centerline to complete the ribbon definition. It is then possible to compute the geometry and the topology of this double-helix ribbon.

**4.2.2. Ideal DNA Ribbon Geometry.** The numerical descriptions of the A-, B-, and C-DNA ribbons given in Table 1 and their three-dimensional representations in Figure 5 capture many well-known features of the canonical double-helical structures. For example, the significantly shorter contour length  $L$  of the A-DNA single-strand ribbon compared to the lengths of the B- and C-DNA ribbons originates from the closer spacing between the phosphorus atoms on either side of the (C3'-endo) puckered sugar rings in the A form compared to the P–P distances across the (C2'-endo) sugar rings in the B and C forms.<sup>51</sup> The contour lengths of the double-helix ribbons are appreciably shorter than those associated with the single strands (roughly half to two-thirds the length). The relative lengths of the double-helix ribbons reflect the different spatial pathways of the paired bases as well as the spacing between adjacent residues in the three structures. Whereas the central core of B DNA describes a nearly straight line, the base pairs attached to the A and C structures follow helical pathways. These differences, which arise from the known lateral displacement, or slide, of successive base pairs in the A and C forms<sup>50</sup>, give rise to the smaller curvature  $\bar{\kappa}$  of the B double-helix ribbon compared to the corresponding A and C ribbons. The differences in curvature are less pronounced for the backbone ribbons, all of which trace out helical pathways.

The integrated twist density  $\theta$  at the vertices of the strand ribbons is roughly half the cylindrical angle  $\Omega_h$ , also called the helical twist,<sup>52</sup> which is commonly used to describe the

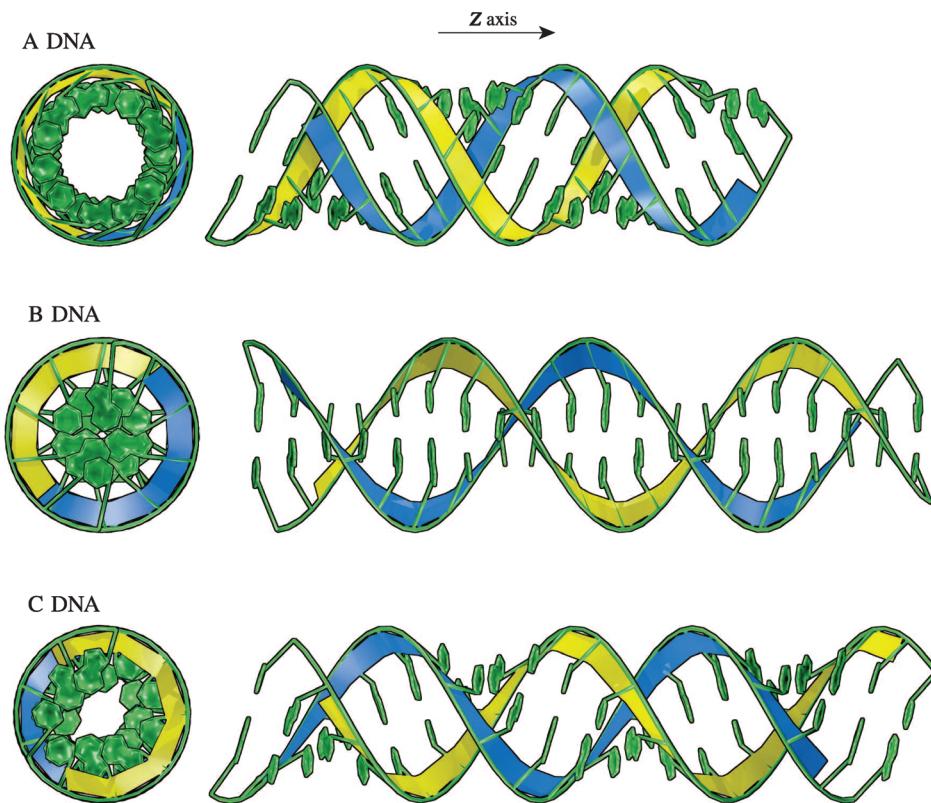
**Table 1. Geometry of Ideal A, B, and C DNA<sup>a</sup>** (see refs 44 and 45)

	A DNA	B DNA	C DNA
Strand Ribbon Quantity			
$n$	11	10	9
$L$ (Å)	103.86	125.88	125.90
$\bar{\kappa}$ (deg·Å <sup>-1</sup> )	5.28	4.65	5.20
$\bar{\tau}$ (deg·Å <sup>-1</sup> )	2.85	2.84	3.13
$\theta$ (°)	15.57	18.80	20.71
$\psi$ (°)	146.54	84.31	60.24
Double-Helix Ribbon Quantity			
$n$	11	10	9
$L$ (Å)	58.46	64.26	83.76
$\bar{\kappa}$ (deg·Å <sup>-1</sup> )	5.90	0.68	5.90
$\bar{\tau}$ (deg·Å <sup>-1</sup> )	8.88	10.62	6.95
$\theta$ (°)	27.31	35.93	30.66
$\psi$ (°)	55.90	93.73	131.14

<sup>a</sup>Ribbons defined either at the strand (top) or at the double-helix (bottom) level for each type of ideal form. The mean value for the curvature  $\bar{\kappa}$  is computed using eq 9, that for the twist density  $\bar{\tau}$  with eq 12, and that for the integrated twist density  $\theta$  with eq 11. The inclination angle  $\psi$  of the ribbon is defined as the angle between any  $\mathbf{m}_1$  vector (see Figure 1) of the discrete ribbon and the double-helical axis of the DNA fragments  $Z = (0,0,1)$ . The contour length  $L$  is the sum of the lengths of the segments along the centerline. The number of base pairs per helical turn  $n$  is a characteristic of each ideal DNA.

orientation of successive residues in each canonical structure with respect to the global helical Z axis, i.e.,  $\Omega_h = 360^\circ/n$ , where  $n$  is the number of nucleotides per helical turn. The integrated twist density of the ribbons constructed for A and C DNA is also a fraction of the helical twist (0.83 and 0.77, respectively), but that of the double-helix ribbon for B DNA is nearly identical to the cylindrical angle. These differences reflect the degree of wrapping of the ribbon centerline about the global helical axis, and, as noted by Britton et al.,<sup>25</sup> also stem from the local chiral distortions associated with the lateral displacement and bending of successive base pairs that produce such pathways. Indeed, the ribbons capture the latter bending through the inclination angle  $\psi$ . Whereas the cross-section of the B-DNA double-helix ribbon is almost perpendicular to the double-helix axis, those of the A and C ribbons are inclined at angles that differ by 35°–40° from the perpendicular orientation. In the case of A DNA, in which the base-pair steps bend in the direction of the major groove,<sup>50</sup> the inclination of the ribbon is less than 90°, and in the case of C DNA, in which the base-pair steps bend in the direction of the minor groove,<sup>50</sup> the inclination angle is greater than 90°. The inclination of the strand ribbons varies in the opposite sense, with that for A DNA being greater than 90° and that for C DNA being less. The differences in strand orientation within individual structures and among the different forms are immediately evident from the relative positions of the backbones and strand ribbons in Figure 5. The twist density  $\bar{\tau}$ , which is given for completeness in Table 1, can be derived, in the case of the ideal helices, from the integrated twist density, the contour length, and the number of base pairs  $N$  in the structural fragment, i.e.,  $\bar{\tau} = (N - 1)\theta/L$ .

**4.3. Application to the Geometry and Topology of a DNA Minicircle.** We next examine a covalently closed, 65 bp double-stranded DNA minicircle, which was captured in recent atomistic molecular dynamics simulations<sup>17</sup> and found to have a linking number  $Lk = 6$  (see Figure 6). The pathway of the



**Figure 5.** Strand ribbon representations of ideal (top) A-, (middle) B-, and (bottom) C-DNA double helices bearing the same  $(ACTGA)_4$  repeating sequence.<sup>44,45</sup> Each ideal structure is depicted by a simplified (base, phosphorus) atomic representation, in which the bases and the vectors that connect the P and N1/N9 atoms are shown in green (images obtained with Mathematica). The discrete ribbons, defined by these atoms, are represented in blue along the sequence-bearing strand and in yellow along the complementary strand. The former strand propagates along the positive Z axis. Note the different lengths, orientations, and global displacements of the ribbons in the three helical forms.

helical axis of the minicircle is nearly planar, and hence its writhing number is close to zero (see Table 3). We first study the geometry and the topology of the minicircle at the strand level, and then we characterize the geometry and topology at the double-helix level.

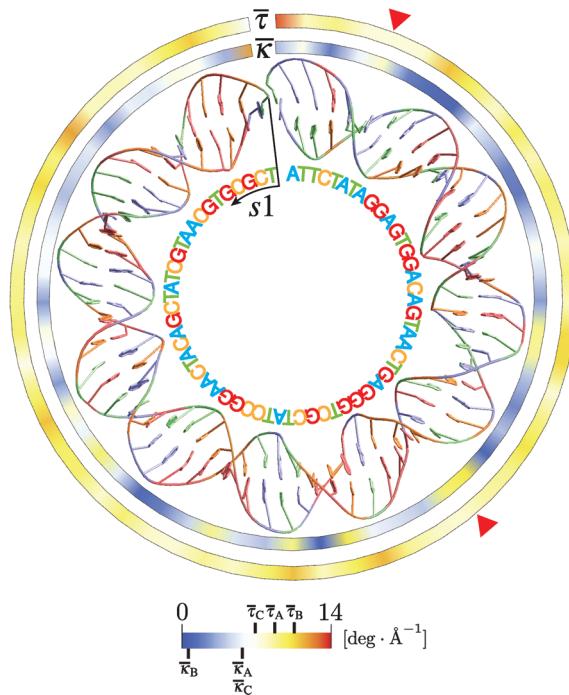
**4.3.1. Strand Level Analysis.** The discrete ribbons for both strands of the minicircle are obtained with the same method used to treat the wrapping of ideal DNA, i.e., derivation of the centerline and material frames from the phosphorus and glycosyl nitrogen atoms. In addition, we also construct ribbon representations for each strand using nitrogen and carbon atoms of the bases. The atoms along the centerline are the glycosyl nitrogens, and the atoms used to derive the material frames are the carbons that lie on the outer, nonhydrogen-bonded edges of the bases (C8 for adenine and guanine and C6 for thymine and cytosine). We can then compare the behavior of ribbons obtained using different choices of atomic coordinates. The results of our analysis are given in Table 2.

**4.3.2. Double-Helix Analysis.** We compute a double-helix ribbon representation of the minicircle using the same procedure applied to ideal DNA fragments. We compare the properties of the ribbon based on the P atoms from hydrogen-bonded residues with the corresponding quantities computed for a ribbon described by the base-pair origins and frames (which are given in terms of the long, short, and normal axes determined with the 3DNA software).<sup>50</sup> The integrated twist density within the base-pair representation has been addressed in the work of Britton et al.<sup>25</sup> as explained in Appendix B. The other quantities can be computed by applying our methods to

the centerline defined by the origins of the base-pair frames. The results are given in Table 3.

**4.3.3. Minicircle Geometry and Topology.** Analysis of the DNA minicircle shows that the linking number is the same regardless of the atoms used to define the discrete ribbon and whether or not the structure is described at the strand or at the double-helix level. The invariance in Lk is easy to understand since the minicircle contains only nonmelted base pairs. That is, the strand ribbons are *contained* in the double-helix ribbon, and hence all the ribbons have the same linking number.

The choice of atom pairs naturally affects the shape of the ribbon and the relative contributions of the writhing number Wr and the total twist Tw to the linking number Lk. For example, the centerline of the strand ribbon described by the glycosyl (N1/N9) nitrogens is a more planar pathway with lesser Wr than that defined by the phosphorus atoms (ribbons respectively labeled (N9/N1, C8/C6) and (P, N9/N1) in Table 2). The writhing number of the double-helix ribbon is even lower and closer to zero, particularly if described by the positions of the base-pair origins (Table 3). Thus, the total twist is greater for a ribbon that passes through the latter points and smaller for a strand-based ribbon with the centerline connecting phosphorus atoms. These compensatory differences also appear in the average integrated twist density  $\theta$ , which varies between  $\sim 18^\circ$  and  $\sim 33^\circ$  in the tabulated examples and in the twist density  $\bar{\theta}$ , which spans the range between  $\sim 2.75 \text{ deg} \cdot \text{\AA}^{-1}$  and  $\sim 9.06 \text{ deg} \cdot \text{\AA}^{-1}$ . The analysis further shows that the sequence-bearing strand s1 is slightly more writhed than the complementary strand s2 regardless of the choice of reference atoms and that the compensatory



**Figure 6.** Sequential variation of the curvature  $\bar{\kappa}$  and the twist density  $\bar{\tau}$ , both in units of degree per Ångström, along the double-helix ribbon constructed from the P atoms of a 65 bp simulated DNA minicircle.<sup>17</sup> The base pairs and P atoms in the simplified atomic representation (obtained with Mathematica) are distinguished by the following sequence color code: A-T blue, T-A green, G-C red, C-G orange. The same color coding is used in the literal sequence of base pairs written inside the minicircle. The arrow denotes the 5'-3' direction of the strand (*s*1) that bears this sequence. The color-coded scale of deformation values, i.e., curvature and twist density values, spans the range [0 deg·Å<sup>-1</sup>; 14 deg·Å<sup>-1</sup>]. The marks above the scale denote colors that correspond to the twist density values of ideal A-, B-, and C-DNA double-helix ribbons, while those below the scale denote colors associated with the curvature of the same ideal DNAs. The red triangles denote regions of high curvature and low twist where melting is likely to occur.

differences in  $Tw$ ,  $\theta$ , and  $\bar{\tau}$  persist. The discrete curvature of strand *s*2, however, consistently exceeds that of strand *s*1.

The variation in twist along the strand ribbon constructed from the nitrogen and carbon base atoms is significantly larger than that along any of the other ribbons. The covalent linkage between the

**Table 3. Topology and Geometry of a Simulated DNA Minicircle<sup>17</sup> at the Double-Helix Level for Two Types of Discrete Ribbons<sup>a</sup>**

quantity	(P, P)	base pairs
Lk	6	6
Wr	0.078	-0.007
Tw	5.922	6.007
<i>L</i> (Å)	239.90	239.63
$\bar{\kappa}$ (deg·Å <sup>-1</sup> )	4.26 (2.40)	5.38 (2.78)
$\bar{\tau}$ (deg·Å <sup>-1</sup> )	9.00 (1.47)	9.06 (1.40)
$\theta$ (°)	32.80 (3.17)	33.27 (4.64)

<sup>a</sup>One ribbon, here denoted (P, P), is constructed from the phosphorus atoms attached to hydrogen-bonded base pairs. The other ribbon is generated from the origins and coordinate frames assigned to DNA base pairs by the 3DNA software. See Tables 1 and 2 for the definitions of the parameters and the equations used to compute numerical values.

selected atom pair, N1–C6 of pyrimidines and N9–C8 of purines, presumably contributes to the wide variation in  $\theta$  and  $\bar{\tau}$ , e.g., deviations of 13–15° in  $\theta$  and roughly 3 deg·Å<sup>-1</sup> in  $\bar{\tau}$ . The variation in  $\theta$  and  $\bar{\tau}$  along a ribbon with the same glycosyl nitrogen centerline but with the material frames derived from the backbone P atoms is even greater (~20° and ~4.5 deg·Å<sup>-1</sup>, respectively). Although the latter ribbon is the mathematical complement of the ribbon constructed using the P atoms to define the centerline and the glycosyl nitrogen atoms to derive the material frames, it is a very different entity. The helical wrapping of the P atoms around the N9/N1 centerline introduces much greater changes in the orientation of material frames. Note that the simple expression that relates  $\theta$  and  $\bar{\tau}$  in an ideal, regularly repeating double helix, no longer holds for the irregular minicircle.

The color-coded circular strips around the molecular image in Figure 6 illustrate the variation in geometry around the minicircle, here the curvature  $\bar{\kappa}$  and the twist density  $\bar{\tau}$  of the double-helix ribbon defined by the P atoms of paired nucleotides. The localized regions of high curvature and low twist are suggestive of sites where the DNA is likely to melt<sup>17</sup> (see the red triangles in Figure 6). The integrated twist density of this ribbon as well as that of the ribbon constructed from the base-pair origins and axes are consistent with the topological properties of the minicircle. A perfectly planar 65 bp DNA minicircle with a linking number of 6 should have an integrated twist density of 33.23° per base pair (6 × 360°/65 bp). The values of 32.80° and 33.27° per base pair found for the

**Table 2. Topology and Geometry of a Simulated DNA Minicircle<sup>17</sup> at the Strand Level for Two Types of Discrete Ribbons<sup>a</sup>**

quantity	<i>s</i> 1 (P, N9/N1)	<i>s</i> 2 (P, N9/N1)	<i>s</i> 1 (N9/N1, C8/C6)	<i>s</i> 2 (N9/N1, C8/C6)
Lk	6	6	6	6
Wr	2.69	2.66	1.16	1.12
Tw	3.31	3.34	4.84	4.88
<i>L</i> (Å)	436.00	436.58	441.89	438.02
<i>I</i> (Å)	6.71 (0.32)	6.72 (0.30)	4.57 (0.37)	4.55 (0.40)
$\bar{\kappa}$ (deg·Å <sup>-1</sup> )	4.66 (1.23)	4.73 (1.32)	5.22 (2.57)	5.23 (2.26)
$\bar{\tau}$ (deg·Å <sup>-1</sup> )	2.76 (0.86)	2.75 (0.82)	5.91 (3.33)	5.97 (2.90)
$\theta$ (°)	18.04 (5.87)	18.52 (5.66)	26.82 (15.11)	27.05 (13.13)

<sup>a</sup>The ribbons denoted (P, N9/N1) have a centerline formed by successive phosphorus P atoms and material frames defined by the glycosyl nitrogen atoms (N9 for A and G, N1 for C and T); those denoted (N9/N1, C8/C6) have a centerline formed by successive glycosyl nitrogens and material frames defined by the C8 atoms of purine and the C6 atoms of pyrimidine bases. The labels *s*1 and *s*2 refer respectively to the sequence-bearing and the complementary strands. Numerical values of the linking number Lk, writhing number Wr, and total twist Tw are obtained using eqs 15, 17, and 19, respectively. The mean value of the length associated with a vertex is reported as *I*. The values of the remaining quantities are computed with the expressions listed in Table 1. The values in parentheses are standard deviations.

respective double-helix ribbons reflect the small nonplanarity of the minicircle.

Comparison of the ribbons described by the minicircle atoms with those associated with ideal double-helical structures shows that the two strands are B-like in terms of the curvature, twist density, and P–P spacing ( $\sim 6.7 \text{ \AA}/\text{bp}$ , obtained from the quotient of the contour length  $L$  and the number of base pairs  $N$ ). Although the distance between successive P atoms never drops to the  $\sim 5.5 \text{ \AA}$  separation characteristic of A DNA, the sequential variation in curvature and twist density includes many states with values characteristic of A and C DNA. Indeed, the conformational excursions along the double-helix ribbon in Figure 6 extend far beyond the limits of the canonical structures. In fact, the curvature is closer to that of A or C DNA and the integrated twist density lies nearly midway between the A- and B-DNA reference states.

**4.3.4. Melted Minicircle.** We complete our analysis of the topology and the geometry of the minicircle with the examination of a different configuration of the same minicircle, i.e., a configuration obtained at a different time during the simulation. Notably, two bases found in a hydrogen-bonded Watson–Crick arrangement in the previously described structure are no longer paired in this configuration. That is, localized melting has occurred in the minicircle during the course of the simulation as can be seen in Figure 7.

This example illustrates how our methods can be applied to such melted structures and, in particular, how the quantities we introduce to describe the geometry and the topology of DNA at the atomic scale capture the melting.

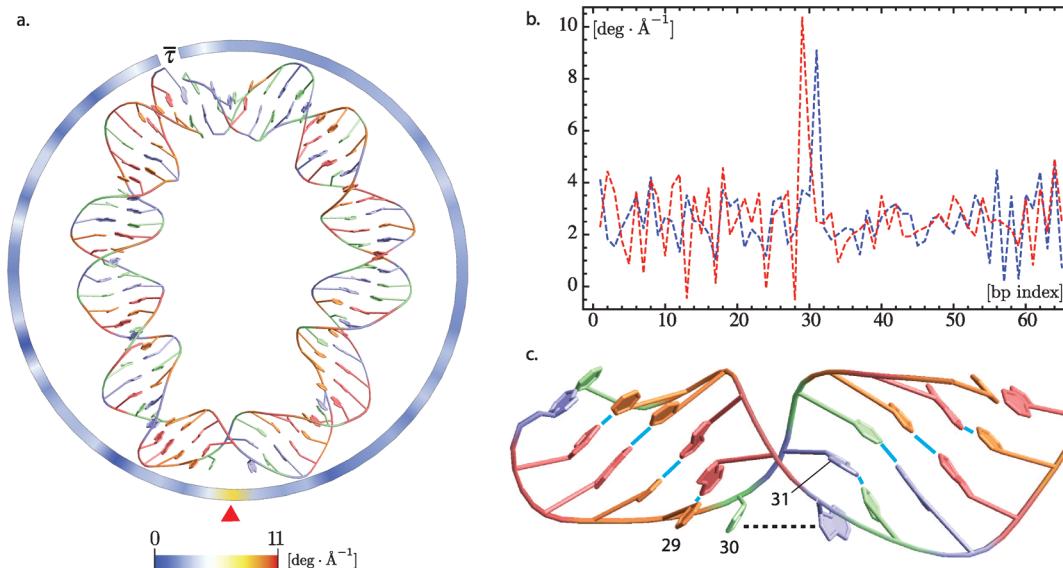
Because the melting in the minicircle is localized to a single base pair, the use of a ribbon representation based on the double helix, i.e., a ribbon built from the phosphorus atoms of both strands, will not capture the changes due to the melting. We need to study two ribbons, one associated with each strand of the minicircle. The strands ribbons are obtained with the same method used to describe the wrapping of ideal DNA at

the strand level, i.e., P and N1/N9 atoms. The results are given in Figure 7.

The melted base pair, depicted by a red arrow in the simplified atomic representation in Figure 7a, is well-captured by the variation in the twist density along the two ribbons obtained for each strand. The twist density in the vicinity of the melted base pair is nearly 5-fold greater than the average value for both strands ( $2.60$  and  $2.65 \text{ deg} \cdot \text{\AA}^{-1}$ , respectively). The strong localization of this large deformation makes it easy to identify the melted site. Moreover, we expect that such a variation in the twist density is a signature of a flipped-out base. In addition, we do not observe significant deviations in the twist density for the strand ribbons from the values obtained for the nonmelted minicircle (given in Table 2) except for the sites where melting occurred.

Although this example shows how our methods can be applied to the treatment of a locally melted structure, it is not intended to describe a general method for the characterization of the topology and the geometry of any melted structure. Indeed, because the minicircle contains only a single melted base pair, the spatial deviations of the residues from the canonical double-helical structure are small. One can expect to have very different results for structures with large melted regions. Moreover, the assessment of the topology of highly melted structures requires a careful examination of the commonly accepted notions of linking number and its decomposition in terms of total twist and writhing number. As pointed out by White and Bauer,<sup>19</sup> the traditional decomposition of the linking number as the sum of the writhe and the total twist does not necessarily hold for melted structures. In the case of more extensively melted structures, one finds instances in which the sum of the total twist and the writhe is an integer other than the linking number. We plan to treat these cases in a forthcoming paper.

**4.4. Application to the Writhing Number Gradient of the Hbb–DNA Complex.** We conclude our series of

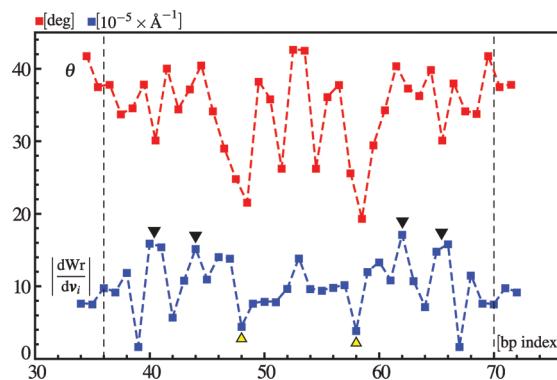


**Figure 7.** Sequential variation of the twist density  $\bar{\tau}$ , in units of degree per Ångström, along the strand ribbons constructed from the phosphorus and glycosyl nitrogen atoms of individual strands of a locally melted 65 bp simulated minicircle. The base pairs and P atoms are depicted (a) as in Figure 6. The color-coded ring around the structure describes the variation in the twist density along the first strand ( $s_1$ ). The line plots (b) show the variation of the twist density along the ribbons for each strand with  $s_1$  in red and  $s_2$  in blue, where  $s_1$  bears the sequence depicted in Figure 6. The melted region is represented (c) in a close-up view, where the numbers denote the base-pair index along  $s_1$ , the blue lines depict paired bases, and the dashed black line indicates the two bases of the melted base pair. Note the sharp bending of DNA about the site of melting.

examples with the computation of the integrated twist density and the writhing number gradient  $dW_r(\{v_i\})/dv_i$  for the S-shaped pseudocontinuous double-helical pathway found in the high-resolution crystal structure of DNA with the Hbb protein from *B. burgdorferi*,<sup>46</sup> the tick-borne bacterium that causes Lyme disease in humans.<sup>53,54</sup> The protein assembles as a homodimer on DNA and bends a 35 bp double-helical stretch by over 180°. The meandering pathway comes from the packing of the assembly in the crystal lattice: The terminal base pairs in each protein-bound duplex stack on the base pairs at the ends of the DNA in two other such complexes.

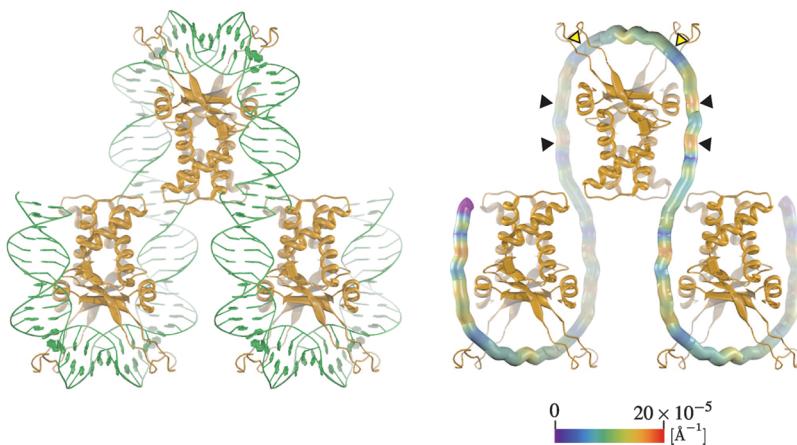
Here we examine a 105 bp DNA fragment (see Figure 8) generated from the atomic coordinates, space group ( $P_{2_1}2_12_1$ ) and cell constants ( $a = 59.27$ ,  $b = 88.32$ ,  $c = 96.57 \text{ \AA}$ ) of the crystal structure (Protein Data Bank entry 2NP2). We proceed along lines similar to those used in the analysis of ideal DNA fragments and the minicircle. We construct a double-helix ribbon from the base-pair frames identified with 3DNA.<sup>50</sup> The centerline connects the base-pair origins, and the material frames emanate from the vertices and coincide with the base-pair axes. We compute the integrated twist density for each segment of the centerline (eq B.1), the values of which are given in Figure 9. We also compute the writhe gradient (eq 25) along the centerline of the ribbon and report its magnitude in Figures 8 and 9. The writhe gradient gives for each vertex the direction along which any moderate local change in the vertex position produces the largest variation in the writhing number. Subsequently, the sites for which the writhe gradient magnitude is the largest correspond to the sites where a small modification of the structure might, all other things being equal or held constant, give rise to noticeable global changes in the topology and the geometry of the DNA structure.

Our characterization of the topology and geometry of the Hbb-bound DNA chain corroborates and complements the spatial features of DNA identified by the crystallographers who determined the structure.<sup>46</sup> For example, the locations of the two large protein-induced bends in the DNA, near the tip of each  $\beta$ -ribbon arm of the protein homodimer (denoted by yellow arrows in Figures 8 and 9), correspond to low values of the integrated twist density and low magnitude of the writhe



**Figure 9.** Integrated twist density ( $\theta$ , in red) and magnitude of the writhe gradient ( $|dW_r/dv_i|$ , in blue) for the Hbb–DNA complex in the middle of the array in Figure 8. The ribbon used to compute the integrated twist density is derived from the base-pair origins and coordinate frames. The latter values of  $\theta$ , in units of degree, are reported for each segment of the ribbon centerline and are computed with the use of eq B.1. See Figure 8 for the computation of the writhe gradient magnitude. The vertical dashed black lines denote the first and last base pairs of the central Hbb–DNA complex (base pair 36 and base pair 70, respectively); the points outside these lines belong to the preceding and succeeding Hbb–DNA complexes. The areas of high and low magnitude of the writhe gradient, denoted respectively by the black and yellow arrows, are discussed in the text.

gradient. Small changes in the base-pair origins at these points would have little effect on the writhing number of a covalently closed DNA that incorporates the Hbb-bound fragment (as long as the DNA ribbon remains free of self-intersection). By contrast, the magnitude of the writhe gradient is nearly 10-fold greater at four sites (denoted by black arrows) where the  $\alpha$ -helical body of the Hbb homodimer contacts the DNA. Small displacement of the base-pair origins at these locations would produce changes in the writhing number of a closed Hbb-bound DNA greater than any that would be brought about by the movement of other base-pair origins. Furthermore, the changes in base-pair position that lead to these large changes in writhe depend upon DNA sequential location. That is, the large changes in  $W_r$  highlighted by the pair of black arrows closer to



**Figure 8.** Simplified atomic representation (left) and magnitude of the writhe gradient (right) for an array of Hbb–DNA complexes.<sup>46</sup> The bases and the P atoms of the depicted 105 bp DNA fragment are represented in green and the backbones of the bound Hbb homodimers in gold. The writhe gradient is computed from the base-pair origins for the same 105 bp fragment by the use of eq 25, and its magnitude, in units of inverse Ångström, is indicated along the color-coded tube. The intensity of colors is proportional to the displacement of atoms and base-pair origins with respect to the plane of the paper, i.e., brightest for points above the plane and dimmest for those below the plane. The black and yellow arrows point respectively to areas of high and low magnitude of the writhe gradient, the numerical values of which are reported in Figure 9 and discussed in the text.

the ends of the homodimer-bound fragment entail motions of the associated base-pair origins out of the plane of Figure 8, while those in the vicinity of the second pair of black arrows arise from motions of the related base-pair origins toward the  $\alpha$ -helices along the horizontal axis of the figure. Whereas the former movements affect the chirality of DNA wrapping around the protein, the latter deformations introduce a kink in the double helical pathway. Finally, it is worth noting that the changes in base-pair positions that would produce greater changes in the writhing number lead to significant changes in the integrated twist density of the two segments of the ribbon that flank the perturbed base-pair origins. These changes in twist occur whether or not the DNA is covalently closed. If the DNA is closed, the variation in the total twist is the equal in magnitude, but of the opposite sign, to the variation in writhing number.

## 5. DISCUSSION

We have presented in this paper a general framework to assess the geometry and the topology of DNA structures in terms of the constituent atoms. This framework is based on a ribbon representation of the DNA, at either the strand or the double-helix level, and makes it possible to compute parameters that capture the geometry and topology associated with different parts of the molecule. Our approach is especially well suited for rapid and accurate characterization of the configurations of large double-stranded structures. We have shown with different examples that our methods can be applied to DNA structures that come from different sources, such as configurations obtained in numerical simulations or structures available in public databases, and that the derived parameters add new insights into the global molecular structures. For example, we can pinpoint those parts of a protein-bound DNA molecule which contribute most to the changes in topology observed experimentally and can identify those points along a covalently closed, circular molecule which would have the greatest effect on its topological properties upon perturbation of the structure. Our methods thus provide a useful tool that complements the traditional analysis of DNA structure at the local level, in terms of quantities such as the rigid body parameters that relate paired bases and the dihedral angles that describe the chemical framework, and introduce a connection between the changes in DNA fine structure and the folding of the molecule as a whole.

The novelty of our work resides, in part, in our capability to quantify the geometry and topology of DNA structures along the pathways described by specific atoms or points in the molecule. We show, in our analyses of three canonical double helices (Section 4.2) and a 65 bp DNA minicircle (Section 4.3), how to construct various ribbon representations of the individual strands. We have found in these examples that the choice of atoms used to define the strand ribbon has a profound effect on the values of the derived geometric and topological parameters. This dependence on the choice of atoms reflects the nonuniformity of the double-helical structure. The spatial pathways taken by the atoms in one part of the molecule (such as base versus backbone atoms) do not necessarily follow those of atoms in another part. The different geometry and topology of ribbons defined by various sets of atoms raise questions about DNA models that are based on point-like representations of the double-helical structure.<sup>11</sup> In order to reproduce known behaviors of the double helix, these models must take correct account of the selected atoms in known DNA structures. For example, the twist of successive base atoms is not necessarily the same as the twist of the

backbone atoms (see Tables 1 and 2). Moreover, it could be interesting to study the relative changes in the geometry and the topology of the atoms forming the DNA backbone versus the atoms forming the bases during the course of molecular dynamics simulations.

Although we have provided examples of the application of our methods to DNA structures, it should be understood that the scope of our methods is broader. The discrete setting we have introduced to describe the geometry and the topology of a ribbon can be readily applied to any molecular structure, such as that of a globular protein, a folded RNA, or a cyclic peptide. For example, one could derive a ribbon representation of a cyclic peptide or folded protein in terms of the coordinates of the  $C\alpha$  and  $C\beta$  atoms in successive amino acid residues. Our methods then make it possible to examine the structures of such molecules in a novel way, such as the identification of distinctly different global configurations in terms of the writhing number without the need for molecular visualization.

Finally, the mathematical framework presented here lays the foundation for quantitative treatment of the topology of more complex DNA structures, such as melted duplexes with flipped-out or unpaired bases, closed circular molecules with extruded cruciforms, multistranded helices, etc. We will show in a forthcoming publication how to take advantage of the strand-based ribbon representation of DNA to address the geometry and the topology of such structures.

The methods described in this paper are implemented in a C++ library for Linux and Unix-based operating systems available upon request to the authors.

## APPENDIX

### A. Segmented Curve Geometry and Topology

The derivation of expressions that deal with certain aspects of the geometry and topology of a segmented curve requires integration across the vertices of such a curve. Here we first introduce a smoothing function that facilitates this integration and then employ the result in integrations to deduce properties of the curvature of the segmented curve and to derive an expression for the small writhe difference,  $dW_r$ , between two similarly shaped, closed, segmented curves by evaluating a linearized Fuller integral, given in its full form in eq 23. Finally, we derive a vector called the writhing number gradient that can be used to give information about the topological properties of closed segmented curves.

**A.1. Segmented Curve Smoothing Procedure.** We consider a discrete closed segmented curve defined by the vertices  $\{\mathbf{v}_i\}$  (see Figure 3). For each vertex we define a relative arc length  $s_i$  such that  $s_i = 0$  corresponds to the vertex position  $\mathbf{v}_i$  and  $-|e^{i-1}|/2 \leq s_i \leq |e^i|/2$ , where  $e^{i-1}$  and  $e^i$  are the edges that meet at the vertex  $\mathbf{v}_i$ . We then specify around each vertex a local domain  $-\varepsilon \leq s_i \leq \varepsilon$ , such that  $\varepsilon \leq |e^{i-1}|/2$  and  $\varepsilon \leq |e^i|/2$ , on which the smoothing of the segmented curve is performed.

We introduce a smoothing function, denoted  $H(s)$ , defined on  $[-\varepsilon; \varepsilon]$ , and assume that  $H(s)$  has the following properties:

- $H(-\varepsilon) = 0$ ,  $H(\varepsilon) = 1$ , and  $H(0) = 1/2$ ,
- $H(s)$  is monotonically smoothly increasing.

We now define on the local domain  $[-\varepsilon; \varepsilon]$  of each vertex a smooth curve with a unit tangent given by

$$\mathbf{t}_i(s) = \frac{(1 - H(s))\mathbf{t}^{i-1} + H(s)\mathbf{t}^i}{\sqrt{1 + 2H(s)(H(s) - 1)(1 - \mathbf{t}^{i-1} \cdot \mathbf{t}^i)}} \quad (\text{A.1})$$

The segmented curve appears in the limit  $\varepsilon \rightarrow 0$ . By setting  $H(s)$  equal to  $1/2$  in eq A.1, we find the tangent at the vertex, i.e.,  $\mathbf{t}_i = (\mathbf{t}^{i-1} + \mathbf{t}^i)/|\mathbf{t}^{i-1} + \mathbf{t}^i| = (\mathbf{t}^{i-1} + \mathbf{t}^i)/(2(1 + \cos \alpha_i))^{1/2}$ , where  $\alpha_i$  is the turning angle at the vertex. It will be seen below that the values of the integrals that we shall be dealing with in our discussion of the curvature and the writhing number, like the tangent at the vertex, are also independent of the value of  $\varepsilon$ , and so the results obtained will be valid for the segmented curves.<sup>55</sup>

**A.2. Average Curvature.** The smoothing procedure makes the segmented curve and its first derivative, the tangent, continuous. Hence, we can express the binormal curvature vector  $\mathbf{k}_i$  on a given vertex domain, as

$$\mathbf{k}_i(s) = \mathbf{t}_i(s) \times \mathbf{t}'_i(s) \quad (\text{A.2})$$

where  $\mathbf{t}'_i(s)$ , the derivative of the tangent, is given by

$$\mathbf{t}'_i(s_i) = \frac{\mathbf{t}^i - \mathbf{t}^{i-1}}{\sqrt{1 + 2H(s_i)(H(s_i) - 1)(1 - \mathbf{t}^{i-1} \cdot \mathbf{t}^i)}} \frac{dH(s_i)}{ds_i} + \quad (\text{A.3})$$

$$\left. \begin{aligned} & ((1 - H(s_i))\mathbf{t}^{i-1} + H(s_i)\mathbf{t}^i) \\ & \times \frac{d}{ds_i} \left( \frac{1}{\sqrt{1 + 2H(s_i)(H(s_i) - 1)(1 - \mathbf{t}^{i-1} \cdot \mathbf{t}^i)}} \right) \end{aligned} \right\} \quad (\text{A.4})$$

The binormal curvature vector, zero everywhere except on the local domain  $[-\varepsilon; \varepsilon]$  of each vertex, is, given eqs A.1 and A.4, equal to

$$\kappa_i(s_i) = \frac{\sin \alpha_i \mathbf{b}_i}{1 + 2H(s_i)(H(s_i) - 1)(1 - \cos \alpha_i)} \frac{dH(s_i)}{ds_i} \quad (\text{A.5})$$

and its magnitude, the curvature  $\kappa_i(s)$ , is

$$\kappa_i(s_i) = \frac{\sin \alpha_i}{1 + 2H(s_i)(H(s_i) - 1)(1 - \cos \alpha_i)} \frac{dH(s_i)}{ds_i} \quad (\text{A.6})$$

Although the curvature depends on the particular choice of a smoothing function, its integral over the vertex domain is independent of the smoothing function. In fact, we see that

$$\begin{aligned} \int_{-\lvert \mathbf{e}^{i-1} \rvert / 2}^{\lvert \mathbf{e}^i \rvert / 2} \kappa_i(s) ds &= \int_{-\varepsilon}^{\varepsilon} \kappa_i(s) ds \\ &= \int_0^1 \frac{\sin \alpha_i}{1 + 2H(s)(H(s) - 1)(1 - \cos \alpha_i)} dH \quad (\text{A.7}) \\ &= \left[ \arctan \left[ (2H - 1) \tan \frac{\alpha_i}{2} \right] \right]_0^1 = \alpha_i \end{aligned}$$

where we used the expression of the curvature in terms of the smoothing function to conclude the calculation, by changing the integration variable from  $s$  to  $H(s)$ . This result is not surprising since the integral of the curvature along a piece of planar curve is always of the form<sup>34</sup>  $\gamma + 2\pi n$ , where  $\gamma$  is the angle between the tangents at both ends of the piece of curve and  $n$  is an integer.

Given eq A.7, we can thus state that the average curvature  $\bar{\kappa}_i$  associated with a vertex  $\mathbf{v}_i$  is

$$\bar{\kappa}_i = \frac{1}{l_i} \int_{-\lvert \mathbf{e}^{i-1} \rvert / 2}^{\lvert \mathbf{e}^i \rvert / 2} \kappa_i(s) ds = \alpha_i / l_i \quad (\text{A.8})$$

**A.3. Linearized Fuller Integral:  $dWr$  and the Writhing Number Gradient.** We consider a segmented closed curve,  $C_1$ ,

and a second such curve  $C_2$ , generated from  $C_1$  by applying infinitesimal changes to the vertex positions. For each segment of  $C_2$  we can then express the unit tangent as

$$\mathbf{t}_2^i = \mathbf{t}_1^i + \delta \mathbf{t}^i \quad (\text{A.9})$$

where  $\delta \mathbf{t}^i$  is the infinitesimal change in the tangent, and  $\mathbf{t}_1^i$  is the unit tangent of the corresponding segment of  $C_1$ . Since  $\mathbf{t}_1^i$  and  $\mathbf{t}_2^i$  are unit tangents, the vector  $\delta \mathbf{t}^i$  must be normal to  $\mathbf{t}_1^i$ . For example, for the  $i$ -th vertex, the change in the tangent  $\mathbf{t}^i$ ,  $\delta \mathbf{t}^i$ , brought about by a small shift  $\mathbf{u}_i$  and  $\mathbf{u}_{i+1}$  in the coordinates of vertices  $\mathbf{v}_i$  and  $\mathbf{v}_{i+1}$  is given to a linear approximation by

$$\delta \mathbf{t}^i = \frac{\mathbf{u}_{i+1} - \mathbf{u}_i}{\lvert \mathbf{e}^i \rvert} - \frac{(\mathbf{u}_{i+1} - \mathbf{u}_i) \cdot \mathbf{t}_1^i}{\lvert \mathbf{e}^i \rvert} \mathbf{t}_1^i \quad (\text{A.10})$$

and it is seen that, as expected,  $\mathbf{t}_1^i \cdot \delta \mathbf{t}^i = 0$ .

We can now apply the Fuller formula eq 23 for the smoothed-out curves  $C_1$  and  $C_2$ , and we can write the first-order approximation of the writhe increment as

$$dWr = - \frac{1}{2\pi} \sum_i \int_{-\lvert \mathbf{e}^{i-1} \rvert / 2}^{\lvert \mathbf{e}^i \rvert / 2} \mathbf{k}_i(s_i) \cdot \delta \mathbf{t}^i = \sum_i \delta Wr_i \quad (\text{A.11})$$

in which  $\delta Wr_i$  is the contribution to  $dWr$  that arises from the integration across the  $i$ -th vertex. Using our previous results for the curvature of a segmented curve we find that

$$\begin{aligned} \delta Wr_i &= - \frac{1}{2\pi} \int_{-\varepsilon}^{\varepsilon} \frac{\kappa_i(s)[(1 - H(s))\delta \mathbf{t}^{i-1} + H(s)\delta \mathbf{t}^i]}{[1 - 2H(s)(1 - H(s))(1 - \cos \alpha_i)]^{3/2}} ds \quad (\text{A.12}) \\ &= - \frac{1}{2\pi} \left( \frac{\mathbf{b}_i \cdot (\mathbf{u}_i - \mathbf{u}_{i-1})}{\lvert \mathbf{e}^{i-1} \rvert} + \frac{\mathbf{b}_i \cdot (\mathbf{u}_{i+1} - \mathbf{u}_i)}{\lvert \mathbf{e}^i \rvert} \right) \tan \frac{\alpha_i}{2} \end{aligned}$$

We can reorganize the terms in the sum and express the writhe increment as a sum of contributions that depend on the change in the position of a given vertex, that is

$$dWr = \sum_i \mathbf{W}_i \cdot \mathbf{u}_i = \sum_i dWr_i \quad (\text{A.13})$$

where the vector  $\mathbf{W}_i$  depends only on the vertex positions, and is defined as

$$\begin{aligned} \mathbf{W}_i &= - \frac{1}{2\pi} \left[ \frac{\mathbf{b}_{i-1}}{\lvert \mathbf{e}^{i-1} \rvert} \tan \frac{\alpha_{i-1}}{2} + \left( \frac{\mathbf{b}_i}{\lvert \mathbf{e}^{i-1} \rvert} - \frac{\mathbf{b}_i}{\lvert \mathbf{e}^i \rvert} \right) \tan \frac{\alpha_i}{2} \right. \\ &\quad \left. - \frac{\mathbf{b}_{i+1}}{\lvert \mathbf{e}^i \rvert} \tan \frac{\alpha_{i+1}}{2} \right] \quad (\text{A.14}) \end{aligned}$$

We thus have an expression for the rate of change of the writhe increment with respect to the vertex positions. In other words, we can write

$$\frac{d}{d\mathbf{v}_i} Wr(\{\mathbf{v}_i\}) = \lim_{\lambda \rightarrow 0} \frac{Wr(\{\mathbf{v}_i + \lambda \mathbf{u}_i\}) - Wr(\{\mathbf{v}_i\})}{\lambda \lvert \mathbf{u}_i \rvert} = \mathbf{W}_i \quad (\text{A.15})$$

which means that the vector  $\mathbf{W}_i$  is the gradient of the writhing number of the ribbon with respect to the vertex position  $\mathbf{v}_i$ .

This expression for the writhe gradient is similar to that previously obtained by de Vries<sup>40</sup> with the use of vertex position interpolation and more involved calculations (we believe that the discrepancy between eq A.14 and the expression in the cited work is a typographical error). Aldinger et al.<sup>41</sup> also obtained similar results but for smooth curves.

## B. Twist Density Across a Segment

We present here the results obtained by Britton et al.<sup>25</sup> related to the calculation of the total twist density for a segmented curve.

We consider two consecutive vertices,  $\mathbf{v}_i$  and  $\mathbf{v}_{i+1}$ , and assume that the material frames emanating from those vertices are adapted. This assumption implies that the tangent at the first vertex is  $\mathbf{t}_i = (\mathbf{t}^{i-1} + \mathbf{t}^i)/\|\mathbf{t}^{i-1} + \mathbf{t}^i\|$  and at the second  $\mathbf{t}_{i+1} = (\mathbf{t}^i + \mathbf{t}^{i+1})/\|\mathbf{t}^i + \mathbf{t}^{i+1}\|$ . The integrated twist density going from the first material frame,  $\mathbf{m}_i$ , to the second one,  $\mathbf{m}_{i+1}$ , can be decomposed as the sum of three distinct contributions, that is

$$\theta^i = \|\mathbf{e}^i\| \bar{\tau}^i = \beta_i + \phi^i + \beta_{i+1} \quad (\text{B.1})$$

where  $\theta^i$  is the integrated twist density associated with a segment (as opposed to  $\theta_i$  associated with a vertex, see (eq 11)). The angles  $\beta_i$  and  $\beta_{i+1}$  are evaluated at each vertex by calculation of the angle between  $\mathbf{m}_{1,i}$  and  $\mathbf{b}_i$  and between  $\mathbf{b}_{i+1}$  and  $\mathbf{m}_{1,i+1}$ , that is

$$\begin{cases} \cos \beta_i = \mathbf{m}_{1,i} \cdot \mathbf{b}_i \\ \sin \beta_i = \mathbf{t}_i \cdot (\mathbf{m}_{1,i} \times \mathbf{b}_i) \end{cases} \quad (\text{B.2})$$

and

$$\begin{cases} \cos \beta_{i+1} = \mathbf{m}_{1,i+1} \cdot \mathbf{b}_{i+1} \\ \sin \beta_{i+1} = \mathbf{t}_{i+1} \cdot (\mathbf{b}_{i+1} \times \mathbf{m}_{1,i+1}) \end{cases} \quad (\text{B.3})$$

Finally, the angle  $\phi^i$  is simply the angle between the two consecutive binormal vectors  $\mathbf{b}_i$  and  $\mathbf{b}_{i+1}$  and is defined by

$$\begin{cases} \cos \phi^i = \mathbf{b}_i \cdot \mathbf{b}_{i+1} \\ \sin \phi^i = \mathbf{t}^i \cdot (\mathbf{b}_i \times \mathbf{b}_{i+1}) \end{cases} \quad (\text{B.4})$$

It also follows from these results that the total twist of a ribbon configuration can be written as

$$\text{Tw} = \frac{1}{2\pi} \sum_i \beta_i + \phi^i + \beta_{i+1} \quad (\text{B.5})$$

where the sum is taken over all the segments of the ribbon centerline.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: clauvelin@biomaps.rutgers.edu; itobias@rci.rutgers.edu.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was generously supported by USPHS Research Grant GM 34809. We are grateful to Sarah Harris and Jonathan Mitchell for bringing to our attention the need for methodology for rapid and accurate assessment of the geometry and the topology of large DNA data sets and for sharing the coordinates of DNA minicircles generated in their molecular dynamics simulations.<sup>17</sup> We also acknowledge Gerald S. Manning for clarification and proof of the Fuller writhe increment formula taken from his private correspondence with F. B. Fuller.

## ■ REFERENCES

- (1) Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163–11168.
- (2) Hud, N. V.; Plavec, J. *Biopolymers* **2003**, *69*, 144–158.
- (3) Farwer, J.; Packer, M. J.; Hunter, C. A. *Biopolymers* **2006**, *81*, 51–61.
- (4) Tolstorukov, M. Y.; Colasanti, A. V.; McCandlish, D. M.; Olson, W. K.; Zhurkin, V. B. *J. Mol. Biol.* **2007**, *371*, 725–738.
- (5) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. *Nature* **1998**, *74*, 773–779.
- (6) Katritch, V.; Vologodskii, A. *Biophys. J.* **1997**, *72*, 1070–1079.
- (7) Merlitz, H.; Rippe, K.; Klenin, K. V.; Langowski, J. *Biophys. J.* **1998**, *74*, 773–779.
- (8) Villa, E.; Balaeff, A.; Schulten, K. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6783–6788.
- (9) Czapla, L.; Swigon, D.; Olson, W. K. *J. Chem. Theory Comput.* **2006**, *2*, 685–695.
- (10) Lankaš, F.; Lavery, R.; Maddocks, J. H. *Structure* **2006**, *14*, 1527–1534.
- (11) Knotts, I. T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. *J. Chem. Phys.* **2007**, *126*, 084901.
- (12) Langowski, J.; Heermann, D. W. *Semin. Cell Dev. Biol.* **2007**, *18*, 659–667.
- (13) Harris, S. A.; Laughton, C. A.; Liverpool, T. B. *Nucleic Acids Res.* **2008**, *36*, 21–29.
- (14) Czapla, L.; Swigon, D.; Olson, W. K. *J. Mol. Biol.* **2008**, *382*, 353–370.
- (15) Arya, G.; Schlick, T. *J. Phys. Chem. A* **2009**, *113*, 4045–4059.
- (16) Geggier, S.; Vologodskii, A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15421–15426.
- (17) Mitchell, J. S.; Laughton, C. A.; Harris, S. A. *Nucleic Acids Res.* **2011**, *39*, 3928–3938.
- (18) Fuller, F. B. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 3557–3561.
- (19) White, J. H.; Bauer, W. R. *J. Mol. Biol.* **1987**, *195*, 205–213.
- (20) White, J. H. *Amer. J. Math.* **1969**, *91*, 693–728.
- (21) Călugăraru, G. *Czech. Math. J.* **1961**, *11*, 588–625.
- (22) Gauss, C. F. In *Carl Friedrich Gauss Werke. Herausgegeben von der K. Gesellschaft der Wissenschaften zu Göttingen*; Schering, E. C. J., Brendel, M., Eds.; Göttingen, Gedruckt in der Dieterichschen Universitätsdruckerei (W.F. Kaestner): Göttingen, Germany, 1867; Vol. V, pp 605.
- (23) Gray, J.; Epple, M. *Math. Intelligencer* **1998**, *20*, 45–52.
- (24) Levitt, M. *J. Mol. Biol.* **1983**, *170*, 723–764.
- (25) Britton, L. A.; Olson, W. K.; Tobias, I. *J. Chem. Phys.* **2009**, *131*, 245101.
- (26) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.; Srinivasan, A.; Schneider, B. *Biophys. J.* **1992**, *63*, 751–759.
- (27) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (28) Langer, J.; Singer, D. A. *SIAM Review* **1996**, *38*, 605–618.
- (29) Bishop, R. L. *Am. Math. Mon.* **1975**, *82*, 246–251.
- (30) Fuller, F. B. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 815–819.
- (31) Pohl, W. *Indiana Univ. Math. J.* **1968**, *17*, 975–985.
- (32) White, J. H.; Bauer, W. R. *J. Mol. Biol.* **1986**, *189*, 329–341.
- (33) Bergou, M.; Wardetzky, M.; Robinson, S.; Audoly, B.; Grinspun, E. Proceedings from ACM SIGGRAPH 2008, Singapore, December 11–13, 2008; ACM: New York, 2008; pp 1–12.
- (34) Grinspun, E. *Introduction to Discrete Differential Geometry: The Geometry of Plane Curves*; ACM: New York, 2006.
- (35) Jeffreys, H. S.; Jeffreys, B. J. *Methods of Mathematical Physics*; Cambridge University Press: Cambridge, UK, 1946.
- (36) Swigon, D.; Coleman, B. D.; Tobias, I. *Biophys. J.* **1998**, *74*, 2515–2530.
- (37) Klenin, K.; Langowski, J. *Biopolymers* **2000**, *54*, 307–317.
- (38) Hannay, J. H. *J. Phys. A* **1998**, *31*, L321.
- (39) Hao, M.-H.; Olson, W. K. *Macromolecules* **1989**, *22*, 3292–3303.
- (40) de Vries, R. J. *Chem. Phys.* **2005**, *122*, 064905.

- (41) Aldinger, J.; Klapper, I.; Tabor, M. *J. Knot. Theory Ramif.* **1995**, 4, 343–372.
- (42) Neukirch, S.; Starostin, E. L. *Phys. Rev. E* **2008**, 78, 041912.
- (43) Vologodskii, A. V.; Cozzarelli, N. R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, 23, 609–643.
- (44) Chandrasekaran, R.; Arnott, S. In *Numerical Data and Functional Relationships in Science and Technology, Group VII, Nucleic Acids*; Saenger, W., Ed.; Springer-Verlag: Berlin, 1989; Vol. 1b, pp 31–170.
- (45) van Dam, L.; Levitt, M. H. *J. Mol. Biol.* **2000**, 304, 541–561.
- (46) Mouw, K. W.; Rice, P. A. *Mol. Microbiol.* **2007**, 63, 1319–1330.
- (47) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, 171, 964–967.
- (48) Franklin, R. E.; Gosling, R. G. *Nature* **1953**, 171, 740–741.
- (49) Marvin, D. A.; Spencer, M.; Wilkins, M. H. F.; Hamilton, L. D. *Nature* **1958**, 182, 387–388.
- (50) Lu, X.-J.; Olson, W. K. *Nucleic Acids Res.* **2003**, 31, 5108–5121.
- (51) Olson, W. K. *Macromolecules* **1975**, 8, 272–275.
- (52) Dickerson, R.; et al. *J. Mol. Biol.* **1989**, 205, 787–791.
- (53) Burgdorfer, W.; Barbour, A.; Hayes, S.; Benach, J.; Grunwaldt, E.; Davis, J. *Science* **1982**, 216, 1317–1319.
- (54) Benach, J. L.; Bosler, E. M.; Hanrahan, J. P.; Coleman, J. L.; Habicht, G. S.; Bast, T. F.; Cameron, D. J.; Ziegler, J. L.; Barbour, A. G.; Burgdorfer, W.; Edelman, R.; Kaslow, R. A. N. *Engl. J. Med.* **1983**, 308, 740–742.
- (55) Cantarella, J. *SINUM* **2005**, 42, 1846–1861.