*Published online 18 May 2011* *Nucleic Acids Research, 2011, Vol. 39, No. 16* ***7289–7299***
*doi:10.1093/nar/gkr260*

# Signatures of DNA flexibility, interactions and sequence-related structural variations in classical X-ray diffraction patterns

**A. A. Kornyshev[1,*], D. J. Lee[1,2], A. Wynveen[3,4] and S. Leikin[5,*]**

[1]Department of Chemistry, Faculty of Natural Sciences, Imperial College London, SW7 2AZ, London, UK, [2]Max Planck Institute for the Physics of Complex Systems, D-01187, Dresden, Germany, [3]School of Physics & Astronomy, University of Minnesota, Minneapolis, MN 55455, USA, [4]Institute for Theoretical Physics II: Soft Matter, Heinrich-Heine-University of Düsseldorf, D-40225 Düsseldorf, Germany and [5]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

Received January 6, 2011; Revised March 29, 2011; Accepted April 7, 2011

## ABSTRACT

**The theory of X-ray diffraction from ideal, rigid helices allowed Watson and Crick to unravel the DNA structure, thereby elucidating functions encoded in it. Yet, as we know now, the DNA double helix is neither ideal nor rigid. Its structure varies with the base pair sequence. Its flexibility leads to thermal fluctuations and allows molecules to adapt their structure to optimize their intermolecular interactions. In addition to the double helix symmetry revealed by Watson and Crick, classical X-ray diffraction patterns of DNA contain information about the flexibility, interactions and sequence-related variations encoded within the helical structure. To extract this information, we have developed a new diffraction theory that accounts for these effects. We show how double helix non-ideality and fluctuations broaden the diffraction peaks. Meridional intensity profiles of the peaks at the first three helical layer lines reveal information about structural adaptation and intermolecular interactions. The meridional width of the fifth layer line peaks is inversely proportional to the helical coherence length that characterizes sequence-related and thermal variations in the double helix structure. Analysis of measured fiber diffraction patterns based on this theory yields important parameters that control DNA structure, packing and function.**

Franklin and Gosling (1) and Wilkins *et al.*'s (2) photographs of X-ray diffraction of DNA fibers have been enshrined in scientific history. Yet, neither have they been fully exploited nor completely understood. Their interpretation within the Watson and Crick model (3) relied on the diffraction theory for fibers of ideal helices (4,5), in which it was assumed that the helices are 'relatively free from the influence of neighboring molecules' (1). These simplifications worked well enough to enable the discovery that DNA is a double helix and for estimating its basic parameters: radius, average helical pitch, average axial rise per base pair and the relative widths of the grooves. Yet these simplifications have precluded extraction of any other information from the photographs. Later studies of oligonucleotide crystals revealed sequence-dependent variations in the double helix structure (6,7), which affect DNA function (8–10). Significant interactions between neighboring molecules in DNA aggregates, which influence the double helix structure (11–13), were observed as well (14). Information about these DNA properties is contained in the X-ray diffraction patterns from non-crystalline aggregates of natural DNA molecules, but an improved diffraction theory is required to extract this information (15).

The diffraction theory for ideal helices has played a key role in the DNA structure discovery that transformed molecular biology. Nevertheless, further development of the theory, to account for non-ideality of the DNA structure, is not just a matter of scientific rigor. We believe that X-ray diffraction from fibers and columnar aggregates of DNA is still a useful tool for better understanding the properties of this remarkable molecule. The information

*To whom correspondence should be addressed. Tel: +44 (0)20 7594 5786; Fax: +44 (0)20 7594 5801; Email: a.kornyshev@imperial.ac.uk
Correspondence may also be addressed to Sergey Leikin. Tel: +1-301-594-8314; Fax: +1-301-402-0292; Email: leikins@mail.nih.gov

contained in the diffraction patterns may help us, e.g. to solve such puzzles as DNA packaging in viruses and cells or homology-based pairing of intact, double-stranded DNA molecules (16,17).

In the present article, we show how X-ray diffraction from oriented non-crystalline aggregates (fiber diffraction) may be utilized for characterizing sequence-related variations in DNA structure and for evaluating the relationship between DNA structure and interactions. We develop a theory for the molecular structure factors of non-ideal, fluctuating double helices confined in aggregates. These factors, which characterize the helical organization of phosphate groups in DNA, determine both the X-ray diffraction patterns and the electrostatic interactions between neighboring double helices (16). We apply this theory for interpreting fiber diffraction patterns and other measurements of *B*-DNA structure, packing and interactions.

In the context of fiber diffraction theory, a step towards modeling non-ideal helices was made in the studies of fibrous protein filaments, in which twist distortions about the major axes of straight molecules were considered (18–20). Based on recent advances in the statistical mechanics of DNA aggregates (21), we utilize a more general approach that incorporates all pertinent sequence-related and thermal fluctuations, large structural defects and effects of intermolecular interactions. A recent renaissance in fiber diffraction studies of various helical macromolecules and assemblies (22) suggests that these developments in the fiber diffraction theory may be important in their own right. Although we focus on DNA, many of our results could be directly used or adapted for the analysis of X-ray diffraction patterns from non-crystalline aggregates of other helical macromolecules.

An improved diffraction theory for fibers and columnar aggregates of DNA forms the foundation for the present study. However, the real focus of this study is on how sequence-related and thermal distortions in the helical coherence (ideality of helical structure) affect properties of nucleic acids and how this information may be deduced from fiber diffraction patterns such as those reported by Franklin and Gosling (1) and Wilkins *et al.* (2). We believe that the corresponding ideas and results may be of interest to a wide audience of researchers interested in nucleic acids rather than only to experts in X-ray diffraction. Therefore, we have tailored the presentation to be accessible to a wider audience. The article is structured accordingly.

In the 'Materials and Methods' section, we provide a qualitative introduction of basic concepts relating the double helix structure to DNA–DNA interactions and DNA packing in aggregates, followed by a general description of X-ray diffraction and molecular structure factors of helical macromolecules. In the 'Results' section, we provide approximations for the molecular structure factors of *B*-DNA that may be utilized for most practical applications. On the basis of these approximations, we re-analyze Franklin and Gosling's diffraction photograph (1) and similar patterns reported later (13,23). We describe what these patterns tell us about the helical coherence of DNA and DNA–DNA interactions. In the 'Discussion' section, we combine the information gleaned from the diffraction patterns with results of other studies, thereby deducing how alignment of strands and grooves on opposing double helices would affect DNA packing in cells and viruses as well as the homologous pairing of DNA. We also consider the role of structural plasticity (adaptation) of DNA in these processes. Since mathematical details of our model are less essential for qualitative analysis of the diffraction patterns and to grasp the implications of this analysis for DNA properties, these details are presented in the Appendix 1. Discussion of the underlying approximations and step by step derivations of all equations are presented online in the Supplementary Data section, for those interested in further mathematical details.

## MATERIALS AND METHODS

### DNA structure, interactions and packing—basic concepts

A crucial feature of DNA packing in fibers and other aggregates—relating DNA structure and interactions—is a zipper-like alignment of strands and grooves with those of neighboring double helices (Figure 1). This alignment was first established in quasi-crystalline DNA fibers at low hydration (24). Similar alignment was later demonstrated also in non-crystalline aggregates at up to ~20 Å surface-to-surface separation between DNA molecules (25). Recently, such an alignment was directly visualized by cryoelectron microscopy of DNA packaged inside viral capsids (26).

Theory suggests that zipper-like strand-groove alignment is an innate feature of interactions between rod-like macromolecules with helical patterns of surface charges (16,27,28). Sugar-phosphate backbone strands of DNA contain a high density of negatively charged, helically arranged phosphates. A significant fraction of these charges is neutralized by counterions, many of which tend to bind in DNA grooves (29). Zipper-like alignment of negatively charged phosphate strands opposite positively charged grooves provides a significant energetic benefit and generates an attractive force (Figure 1C), contributing to the total DNA–DNA interaction (28). Thermal undulations of DNA in hydrated aggregates strengthen this relatively short-ranged electrostatic zipper attraction by bringing parts of the molecules closer together (Figure 1A), extending the interaction range beyond 20–30 Å of surface separation (21).

An important feature of the alignment is that it is affected by sequence-related and thermal fluctuations in the twist and rise (Figure 1B) between adjacent base pairs (16,30). In an isolated molecule, the resulting distortions in the helical periodicity of DNA accumulate with a characteristic length $\lambda_c$ [~30 bp (15)], defined as the helical coherence length of DNA (30) (Figure 2A and B). This length is an important cumulative statistical parameter that characterizes the sequence-dependent structure of DNA (15).

Without compensating deformations, the accumulation of distortions in the helical periodicity would prevent a favorable alignment of molecules with uncorrelated

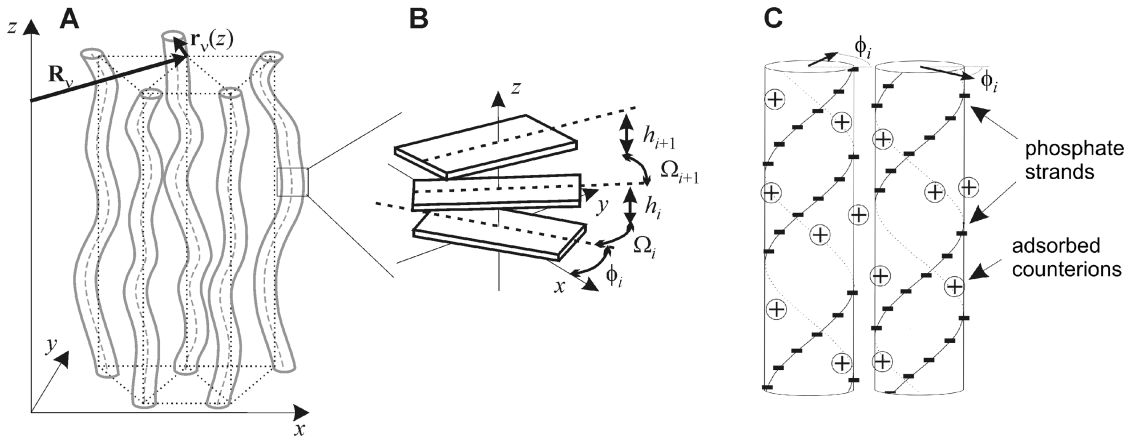


**Figure 1.** Geometry and alignment of DNA molecules in hydrated fibers. (**A**) A cartoon representation of undulating DNA packed in a hexagonal lattice. Dotted straight lines show the hexagonal lattice. Vertical dotted lines represent average positions of DNA centerlines. Dashed lines represent DNA centerlines at a given moment in time. $\mathbf{R}_v$ is the lateral $(x,y)$ coordinate of the average position of the centerline of molecule v; $\mathbf{r}_v$ is the lateral displacement of the centerline from this position. (**B**) Base pair stacking in DNA; each domino represents a base pair; $h_i$ and $\Omega_i$ are the axial rise and twist between base pairs $i$ and $i-1$; $\phi_i$ is the azimuthal orientation of the base pair $i$ (vector pointing towards the middle of the minor groove) defined with respect to the $x$-axis. (**C**) Energetically favorable zipper-like alignment of negatively charged strands across positively charged groove on the neighbor molecule.

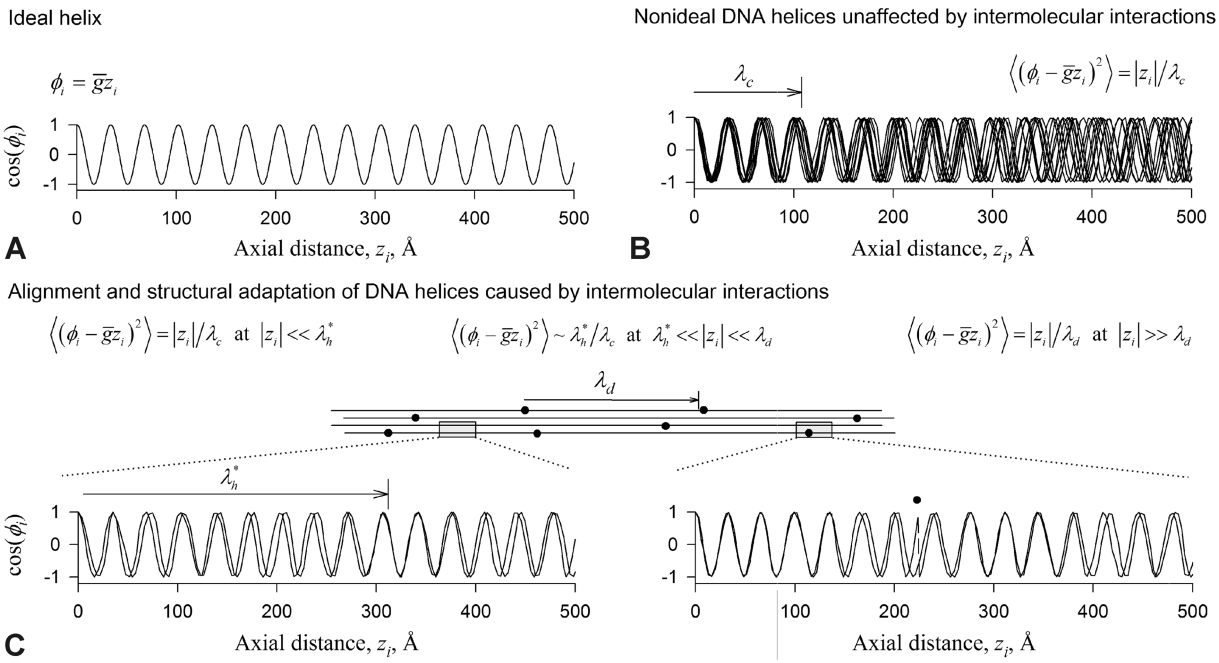


**Figure 2.** Helical coherence of DNA. (**A**) Azimuthal orientations $\phi_i$ of base pairs $i$ in an ideal helix. The axial rise and twist per base pair (c.f., Figure 1B) in an ideal helix are both constant and $\phi_i = \bar{g}z_i$, where $z_i$ is the axial coordinate of the base pair $i$ and $\bar{g}$ is the reciprocal pitch of the helix. We set $z_0 = 0$ and $\phi_0 = 0$. (**B**) Azimuthal orientations of base pairs in 10 DNA molecules with uncorrelated sequences in the absence of structural adaptation. Stacking of base pairs in a real DNA molecule is akin to a walk with a random length of the step; the twist and rise per base pair at each specific step vary around their average values depending on the sequence and thermal motions. When unaffected by other molecules, such stacking leads to a random walk-like deviation of $\phi_i$ from $\bar{g}z_i$, $\left\langle (\phi_i - \bar{g}z_i)^2 \right\rangle \approx |z_i|/\lambda_c$ (15,31). The helical coherence length $\lambda_c$ is a cumulative statistical parameter that describes deviations of DNA structure from that of an ideal double helix with the same average reciprocal pitch. When unaffected by intermolecular interactions, DNAs with different sequences completely lose their alignment at $z_i > \lambda_c$. (C) Schematic representation of a long fiber (middle) and azimuthal orientations of base pairs in neighboring molecules within regions without (lower left) and with (lower right) large structural defects (black dots), such as mismatched bases, flipped-out bases, breaks in the sugar-phosphate backbone, etc. DNA–DNA interactions favor an apposition of negatively charged phosphate strands against positively charged grooves (Figure 1C). At close separations, the energetic benefit of this alignment is sufficient to cause DNA backbone deformation (structural adaptation), preventing unlimited accumulation of deviations from the ideal helical structure at large $z_i$. At juxtaposition lengths larger than the structural adaptation length $\lambda_h^*$, the adaptation makes the helices more ideal and restores their alignment (lower left). For infinitely long molecules without large structural defects, one expects $\left\langle (\phi_i - \bar{g}z_i)^2 \right\rangle \approx |z_i|/\lambda_c$ at $z_i << \lambda_h^*$ and $\left\langle (\phi_i - \bar{g}z_i)^2 \right\rangle \approx \lambda_h^*/\lambda_c$ at $z_i >> \lambda_h^*$ (21,31). For DNA, one expects rare ($\sim$1 per $10^3$–$10^4$ bp), randomly distributed, large structural defects (black dots) to disrupt the helical coherence of individual molecules while resetting the alignment of the neighbors (lower right), resulting in $\left\langle (\phi_i - \bar{g}z_i)^2 \right\rangle \approx |z_i|/\lambda_d$ at $z_i >> \lambda_d$. The characteristic length $\lambda_d$ is determined by the distance between such defects.

sequences at juxtaposition lengths larger than $\lambda_c$ (Figure 2B), thus eliminating the zipper attraction (31). However, when the molecules are separated by <20–30 Å of water, the benefit of the zipper attraction outweighs the cost of deformations needed to restore the alignment (21,31). The resulting structural deformations (adaptation) of DNA alter the accumulation of distortions in the helical periodicity of DNA (Figure 2C). In this case, the distortions accumulate with the characteristic length $\lambda_c$ only at axial length scales smaller than the structural adaptation length $\lambda_h^*$. At larger length scales, the distortions accumulate with a much larger characteristic length $\lambda_d$ ($\lambda_d >> \lambda_c$) [Figure 2C and Appendix 1, Equation (A10)]. The parameter $\lambda_h^*$ is determined by the strength of DNA–DNA interactions and depends on the separation between molecules in the aggregate. A theoretical model for calculating $\lambda_h^*$ was described in Ref. (21). The length scale $\lambda_d$ is determined by the length of DNA molecules and large structural defects in DNA (Figure 2C). The theoretical concept of the structural adaptation of DNA in aggregates (21,31) is supported by observations of the alignment in hydrated DNA assemblies (25,26) and changes in DNA structure upon aggregation (11–13,32,33).

In the present study, we show how X-ray fiber diffraction may be utilized for measurement of the helical coherence length and suggest how it may be used to further test theoretical predictions for the structural adaptation.

### X-ray diffraction from columnar assemblies and fibers—basic concepts

The intensity of X-ray diffraction from an assembly (e.g. a fiber) of parallel helices oriented along the $z$-axis (Figure 1) is a sum of scattering intensities at layer lines $n$ (4,5,25)

$$I(\mathbf{k}) = \sum_n I_n(\mathbf{k}), \tag{1}$$

$$I_n(\mathbf{k}) \approx I_n^M(\mathbf{k}) N_M + I_n^M(\mathbf{k}) \sum_{\nu \neq \mu} \left\langle e^{in\left(\Phi_\mu(z)-\Phi_\nu(z)\right)+i\mathbf{K}\left(\mathbf{R}_\nu+\mathbf{r}_\nu(z)-\mathbf{R}_\mu-\mathbf{r}_\mu(z)\right)} \right\rangle. \tag{2}$$

Here $\mathbf{k} \equiv (k_z, \mathbf{K})$ is the scattering vector with $k_z$ parallel and $\mathbf{K}$ perpendicular to the fiber axis ($z$-axis); $N_M$ is the number of molecules in the X-ray beam; $I_n^M(\mathbf{k})$ is the scattering intensity from one molecule; angular brackets indicate statistical averaging; $\Phi_\nu$ (z) is the helical phase of molecule $\nu$ [Equation (A6)]; and $\mathbf{R}_\nu + \mathbf{r}_\nu(z)$ is the lateral coordinate of the centerline of molecule $\nu$ at the axial position $z$ (Figure 1A).

The first term in Equation (2) describes scattering from individual molecules. The second one accounts for interference between X-rays scattered from different molecules (intermolecular scattering). In hydrated DNA aggregates, intermolecular scattering may be significant at the equator ($n = 0$) and the first two layer lines ($n = \pm 1, \pm 2$), while scattering on individual molecules determines the diffraction intensity at higher layer lines (25).

In DNA assemblies, X-rays are scattered primarily by phosphate groups and $I_n^M(\mathbf{k})$ may be approximated with (16)

$$I_n^M(\mathbf{k}) \approx \frac{f_p^2}{2\pi} s(k_z, K, n) J_n^2(Ka). \tag{3}$$

Here $f_p$ is the scattering amplitude of a phosphate group, $J_n(x)$ is a Bessel function of order $n$ and $a$ is the distance from phosphate centers to DNA centerline (hereafter referred to as the DNA radius). The molecular structure factor of phosphates $s(k_z,K,n)$ is a Fourier–Bessel transform of the phosphate density [Equations (A1,A2)].

For an 'ideal' double helix with a given azimuthal half-width of the minor groove $\tilde{\phi}_s$ ($\approx 0.38\pi$ for $B$-DNA) and $N_p$ phosphates within each of the two strands (16)

$$s(k_z, K, n) = \frac{N_p^2}{\pi^2} \left[\cos n\tilde{\phi}_s\right]^2 \sum_{j=-\infty}^{\infty} \delta_{k_z,\, 2\pi j/\bar{h} - n\bar{g}}, \tag{4}$$

where $\delta_{x,y}$ is the Kroenecker delta ($\delta_{x,y} = 1$ at $x = y$ and $\delta_{x,y} = 0$ at $x \neq y$) and $\bar{g} = \bar{\Omega}/\bar{h}$ is the reciprocal pitch of DNA; $\bar{\Omega}$ and $\bar{h}$ are the twist and axial rise between adjoining base pairs, respectively (Figure 1B). The X-ray scattering amplitudes are therefore non-zero only at layer lines for which (4,5)

$$k_z = 2\pi j/\bar{h} - n\bar{g} \ \ (n, j = 0, \pm 1, \pm 2, \ldots). \tag{5}$$

These equations represent a simplified version of Crick's diffraction theory for ideal helices, which was the key to discovering the double helical structure of DNA (3–5). They describe positions of the diffraction peaks that are determined by the average pitch, radius and relative widths of DNA grooves (Figure 3). However, equations derived for point-like scatterers positioned on ideal double helices predict sharp layer-lines with zero width. To describe the actual width and dependence of the diffraction intensity on $k_z$ at these layer lines, one must account for distortions in the double helix structure (34).

Our theory [Equation (A5)] provides a general relationship between molecular structure factors and pair correlation functions for distortions in the structure of helical macromolecules confined in columnar assemblies. This theory may thus be used to describe the experimentally observed broadening of the layer lines. We focus below on $B$-DNA, for which approximate expressions that are not much more complicated than Equation (4) may be derived, and discuss the layer line broadening observed in the classical $B$-DNA diffraction patterns.

## RESULTS

### Molecular structure factors of $B$-DNA

Similar to ideal double helices, assemblies of real $B$-DNA molecules are expected to produce X-ray diffraction peaks centered at the layer lines with $n = \pm 1, \pm 2, \pm 3$ and $\pm 5$ and no diffraction peaks with $n = \pm 4$ (Figure 3). However, distortions in the double helix structure of $B$-DNA molecules lead to broadening of these peaks.

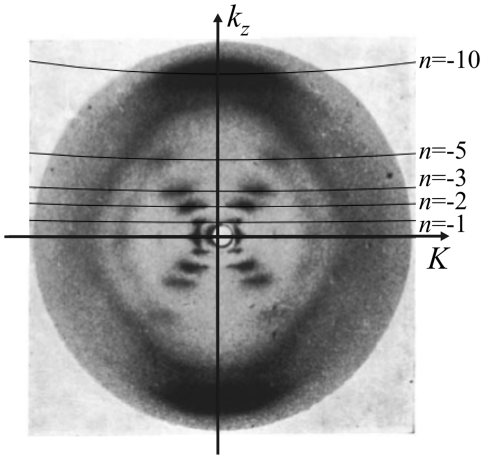


**Figure 3.** A new look at the B-DNA X-ray diffraction pattern photographed by Franklin and Gosling [reproduced with permission from (1)]. Thin solid lines show the layer lines with $k_z = -n\bar{g}$. One can see diffraction peaks centered at $k_z = 2\pi j/\bar{h} - n\bar{g}$ ($j = 0$; $n = 0, \pm 1, \pm 2, \pm 3, \pm 5$ and $j = \pm 1$, $n = 0$). $K$-coordinates of these peaks correspond to the maxima of the Bessel functions $J_n^2(Ka)$. The absence of diffraction peaks at $j = 0$, $n = \pm 4$ suggests that $\cos^2(4\tilde{\phi}_s) \approx 0$. This pattern is consistent with the one expected for a double helix with a radius $a \approx 10$ Å, a helical pitch $2\pi/\bar{g} \approx 34$ Å, $2\pi/\bar{g}\bar{h} \approx 10$ base pairs per helical turn and an azimuthal half-width $\tilde{\phi}_s \approx 0.38\pi$ of the minor groove [Equations (1–8)]. Positions of the diffraction peaks may be explained within the ideal double-helix model [Equation (4)], but their meridional ($k_z$) broadening cannot be attributed just to experimental factors, suggesting a significant contribution from the non-ideal helical structure of DNA [Equations (6–8)]. Indeed, the X-ray beam width should have a similar broadening effect on all layer lines. Its contribution to the diffraction peaks with $n = \pm 3$ and $\pm 5$ should be negligible, because the latter peaks are much broader than the peaks with $n = \pm 1$. The non-linear response of X-ray films might result in additional broadening of the more intense $n = \pm 1, \pm 2, \pm 3$ peaks, but not the $n = \pm 5$ peaks. The broadening of the $n = \pm 5$ peaks is also inconsistent with smearing along concentric arcs due to an imperfect vertical alignment of the molecules in the fibers. The contribution of this smearing, which was estimated based on the arcing of the $n$, $j = 0$ peaks, appears to be small and is consistent with the expected effect of thermal undulations of DNA (Figures 1 and 5). Thus, $kz$-cross-sections of the $n = \pm 5$ peaks in the Franklin and Gosling pattern may be compared with the predictions of our model [Equations (7,8)], while unknown contributions of the x-ray beam width, film nonlinearity and imperfect DNA alignment may significantly affect the other peaks. This analysis yields information about the helical coherence of DNA, as discussed in the text and illustrated in Figure 5. Future studies of structural adaptation of DNA from $k_z$-cross-sections of the other peaks might also be possible, provided that the contribution of experimental factors can be reduced or deconvoluted from the effects of sequence-related and thermal variations in the structure of the double helix.

This broadening can be described by replacing the delta functions in Equation (4) with Lorentzians (Figure 4).

At $n = \pm 1, \pm 2$, one may approximate the molecular structure factors of *B*-DNA by a sum of two Lorentzians (see Appendix 1 and Supplementary Data):

$$s(k_z, K, n) \approx \frac{2N_p}{\pi^2 \bar{h}} \left[\cos n\tilde{\phi}_s\right]^2 \exp\left(-\frac{K^2 d^2}{2}\right) \times \left\{\left(1 - e^{-n^2\lambda_h^*/2\lambda_c}\right) \frac{n^2/2\lambda_c}{(k_z + n\bar{g})^2 + n^4/4\lambda_c^2} + e^{-n^2\lambda_h^*/2\lambda_c} \frac{n^2/2\lambda_d}{(k_z + n\bar{g})^2 + n^4/4\lambda_d^2}\right\} \quad (6)$$

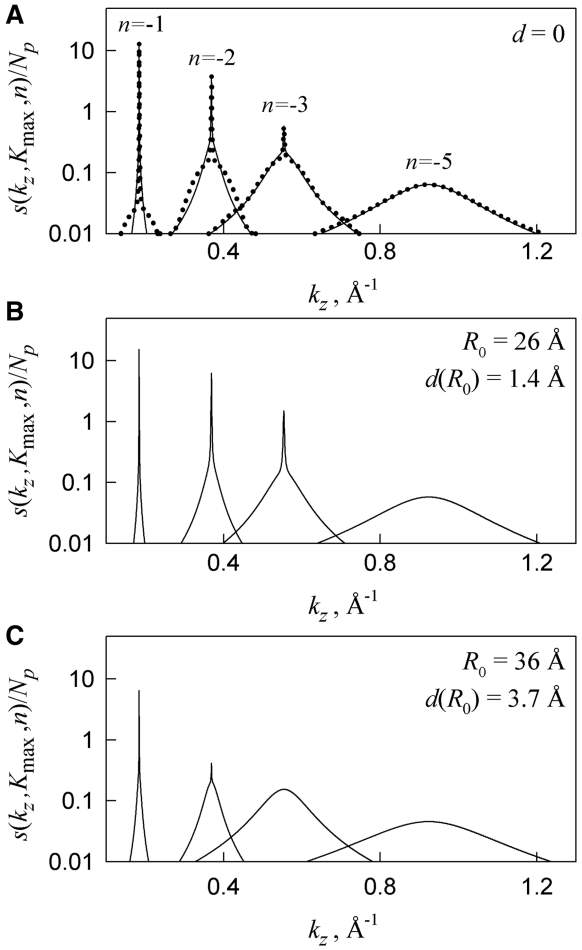


**Figure 4.** Calculated meridional profiles of X-ray diffraction peaks: the plots show the dependence of DNA structure factors on $k_z$ for the first four non-vanishing layer lines ($n = \pm 1, \pm 2, \pm 3, \pm 5$) calculated at $K = K_{max}(n)$ corresponding to the maximum of $[J_n(Ka)]^2$ [see Equation (3)]. (**A**) Comparison of the structure factors calculated for a fiber composed of straight double helices within the approximation given by Equation (6) (solid lines) and from the more general Equations (S22,S33) derived in Supplementary Data (bold dotted lines). The structure factors may be approximated by the sum of sharp and broad Lorentzians [Equation (6)]. The broad one is associated with deviations from the ideal helical structure at axial length scales shorter than the characteristic length $\lambda_h^*$, at which DNA structure adapts to optimize intermolecular interactions in the fiber (here $\lambda_h^* \approx 70$ Å). The width of this Lorentzian, $n^2/2\lambda_c$, is determined by the helical coherence length ($\lambda_c$) of DNA. The sharp Lorentzian originates from axial length scales longer than $\lambda_h^*$, at which the helical periodicity becomes more ideal. Its width $n^2/2\lambda_d$ is determined by the characteristic distance between structural defects in DNA ($\lambda_d$, $\lambda_d >> \lambda_c$). The contribution of this Lorentzian to the structure factor exponentially decreases with $n^2$. (**B** and **C**) Dependence of the structure factors on the average interaxial distance $R_0$ between undulating DNA molecules in hydrated fibers. Based on the criteria given in the main text, the structure factors at $R_0 = 26$ Å were calculated from Equation (6) ($n = -1, -2, -3$) and Equation (7) ($n = -5$). At $R_0 = 36$ Å, we used Equation (6) for the structure factors with $n = -1, -2$ and Equation (7) for the ones with $n = -3, -5$. The transition between the two approximations for the structure factor with $n = -3$ occurs at $R_0 \approx 28$ Å (see main text). Consistently the two approximations produced closely matching $n = -3$ structure factors at $R_0 = 28$ Å (data not shown). The values of $\lambda_h^*$ ($R_0$) for these two panels were calculated as described in Ref. (21). The other parameters were (see main text and Figure 5) $\bar{h} = 3.4$ Å, $\bar{g} = 0.185$ Å$^{-1}$, $\tilde{\phi}_s = 0.38\pi$, $\lambda_c = 105$ Å, $\lambda_d = 1000$ $\bar{h}$, $d(R_0) = 0.225(R_0 - 19.7)$.

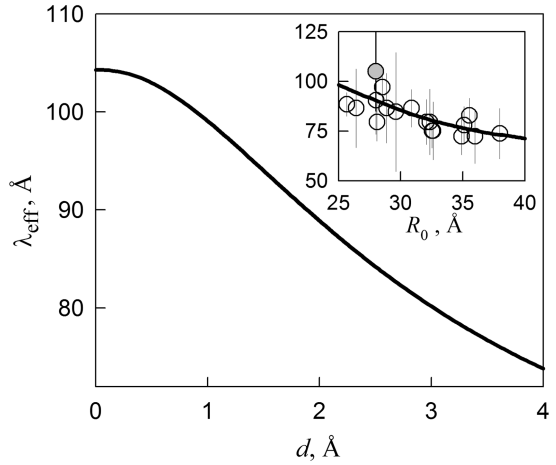


**Figure 5.** Evaluation of the helical coherence length $\lambda_c$ from meridional intensity profiles of X-ray diffraction peaks at the fifth layer line ($n = \pm 5$). The expected profile is a Lorentzian function with the width $n^2/2\lambda_{\text{eff}}$ [Equation (7)]. In the absence of DNA undulations, $\lambda_{\text{eff}} = \lambda_c$. The main plot shows the dependence of $\lambda_{\text{eff}}$ on the undulation amplitude calculated from Equations (8) and (A12). The inset shows $\lambda_{\text{eff}}$ obtained by fitting measured $k_z$ profiles of X-ray diffraction peaks at the fifth layer line with Equation (7) (experimental data points) and $\lambda_{\text{eff}}$ calculated from Equations (8) and (A12) (solid line) at $d(R_0) \approx 0.225(R_0 - 19.7\text{Å})$ (see main text). Original X-ray films with diffraction patterns at different interaxial spacings $R_0$ were generously provided by Zimmermann (13,23) and analyzed as described in Refs. (15,25). The grey data point was obtained by digitizing the diffraction pattern reproduced in Figure 3 from Franklin and Gosling paper (1). This point might be less accurate, because we digitized the published pattern rather than the original X-ray film.

Here $d$ is the root mean square amplitude of the lateral displacement of the DNA centerline due to thermal undulations [Figure 1A, Equation (A7)] and $\bar{h}$ and $\bar{g}$ are the average axial rise per base pair and reciprocal pitch of DNA. The width of the first Lorentzian is determined by the helical coherence length of DNA, $\lambda_c$. The width of the second one is determined by the characteristic length associated with large structural defects in the double helix, $\lambda_d$. The relative weights of the two Lorentzians are determined by the structural adaptation length of DNA, $\lambda_h^*$. These three lengths are defined above in the 'Materials and Methods' section and are graphically illustrated in Figure 2.

At $n = \pm 5$, a single Lorentzian approximation might be used (Supplementary Data)

$$s(k_z, K, n) \approx w_0(K, n, d)\frac{2N_p}{\pi^2\bar{h}}\left[\cos^2\left(n\tilde{\phi}_s\right)\right]\frac{n^2/2\lambda_{eff}}{(k_z+n\bar{g})^2+n^4/4\lambda_{eff}^2}, \quad (7)$$

where

$$\lambda_{eff} = \lambda_c w_1(K, n, d). \quad (8)$$

Here $w_0(K,n,d)$ describes changes in the amplitude and $w_1(K,n,d)$ ($\leq 1$) describes broadening of the diffraction peaks associated with thermal undulations of the helices. At small undulations ($d \to 0$), $w_0(K,n,d) \approx w_1(K,n,d) \approx 1$ and Equations (7) and (8) reduce to Equation (6), in which $\exp(-K^2d^2/2) \approx 1$ and $\exp(-n^2\lambda_h^*/2\lambda_c) \approx 0$. Expressions for $w_0(K,n,d)$ and $w_1(K,n,d)$ are provided in Appendix 1 [Equations (A11,A12)]. The dependence of $\lambda_{eff}$ on $d$ is shown in Figure 5 at values of $K$ corresponding to the maxima of $J_n^2(Ka)$. Because DNA undulations are small at small interaxial distances, $R_0$, between neighboring molecules, Equation (7) with $w_0(K,n,d) \approx w_1(K,n,d) \approx 1$ may be used, e.g. for *B*-DNA fibers at low hydration.

At $n = \pm 3$, one may use Equation (6) at small $R_0$ between DNA and Equation (7) at large $R_0$ (Appendix 1). The exact range for each approximation may depend on DNA base pair composition and sequence [expected to affect $\lambda_c$ (15)], as well as on the concentration and composition of the electrolyte in the aqueous solution in which the molecules lie [expected to affect $\lambda_h^*$ (21)]. For instance, in fibers composed of non-specific natural DNA sequences [$\lambda_c \sim 100$ Å (15)] hydrated by a solution at physiological salt concentrations (0.1–0.2 M NaCl), we expect the transition between the two approximations at $R_0 \approx 28$ Å (Appendix 1).

## *B*-DNA fiber diffraction

To illustrate applications of this model, consider the classical fiber diffraction pattern reported by Franklin and Gosling (1) (Figure 3). As noted above, the diffraction peaks occur at the same positions as expected for ideal helices, which is why Watson and Crick were able to interpret Figure 3 as double helix scattering (3). However, the $k_z$-cross-sections of these peaks are not like those expected for ideal double helices.

For ideal double helices, the only sources providing a finite width for the $k_z$-cross-sections are experimental factors, such as the X-ray beam width, an imperfect alignment (vertical orientation) of the molecules in the fiber and the non-linear response of X-ray films. These factors may contribute to the observed broadening of the peaks, but they cannot fully explain this broadening (Figure 3). Broadening of these cross-sections, associated with distortions in the double helix structure of real DNA molecules, contributes significantly to the diffraction pattern. It depends on the helical coherence length ($\lambda_c$), structural adaptation length ($\lambda_h^*$) and undulation amplitude ($d$) of the double helix, and so broadening reveals information concerning DNA structure and interactions.

Our model predicts $k_z$-cross-sections of the diffraction peaks at the first two layer lines ($n = \pm 1, \pm 2$) to be the sum of broad and narrow Lorentzians [Equation (6), Figure 4], although the sharpness of the second Lorentzian might be exaggerated by the approximations intrinsic to Equation (A10). The predicted narrowing of the diffraction peak at the center of these layer lines is the hallmark of structural adaptation. The relative amplitude of the narrow Lorentzian is determined by $\lambda_h^*$ and represents the extent of the adaptation. Provided that the peak broadening associated with the experimental factors discussed above can be sufficiently reduced, measurements of these $k_z$-cross-sections may be utilized to characterize the extent of structural adaptation experimentally.

At the fifth layer line ($n = \pm 5$), the $k_z$-cross-section may be approximated with a single Lorentzian [Equation (7)].

The width of this Lorentzian is determined by the helical coherence length and undulation amplitude of DNA and it is not affected by structural adaptation [Equations (7,8)]. (Structural adaptation of DNA does not affect this layer line because $n = \pm 5$ helical harmonics of phosphate density do not affect intermolecular interactions and *vice versa*.) It is proportional to $n^2$ and generally large enough, such that experimental artifacts would be negligible (e.g. see Figure 3). This interpretation is supported by the agreement between the widths measured in Ref. (15) and the width calculated from Equations (7,8,A12) without any adjustable parameters (Figure 5, inset). The width of the diffraction peaks at the fifth layer line may therefore be used for experimental measurement of $\lambda_c$ from fiber diffraction. Because the effect of undulations on this width is relatively weak (Figure 5), a fitting of the experimental data in NaCl solutions may be performed, e.g. using $d \approx 0.225(R_0\text{-}19.7\text{Å})$ , where $R_0$ is the interaxial distance between neighboring molecules in the fiber. [This simple approximation for $d(R_0)$ was obtained by linear regression of $d(R_0)$ calculated for DNA in 0.15 M NaCl as described in Ref. (21), but it remains reasonably accurate for NaCl concentrations ranging from 0.05 to 0.3 M.] Alternatively $\lambda_c$ may be estimated from the data at small $R_0$ by assuming $\lambda_c \approx \lambda_{\text{eff}}$. This value may be further corrected for imperfect vertical alignment of molecules in fibers (e.g. due to undulations) based on the observed arcing of different diffraction peaks, as discussed in Ref. (15).

Interpretation of the $k_z$-cross-sections at the third layer line ($n = \pm 3$) may depend on the specific experimental conditions, but they may also provide useful information about $\lambda_c$ and $\lambda_h^*$, as noted in the preceding subsection.

## DISCUSSION

Analysis of the X-ray photograph taken by Franklin and Gosling (1) and similar patterns reported later (13,23) illustrates why diffraction of DNA fibers should not be relegated to scientific history. Combined with the improved theory, it may also be a valuable tool for the future. Fiber diffraction does not provide atomic resolution, but it is not limited to short (10–20 bp), synthetic molecules, like studies of oligonucleotides by X-rays in crystals and by NMR in solution. Atomic resolution structures of longer DNA fragments are available for crystals of protein–DNA complexes, but the structures of these fragments might be significantly affected by protein–DNA interactions and crystal-packing forces. Fiber diffraction reveals structural parameters for long, natural DNA sequences not readily accessible by other techniques, which may be useful for elucidating effects of the double helix structure on the functional properties of DNA. For instance, as illustrated by Figure 5, fiber diffraction may be utilized for measuring helical coherence lengths for sequences with different base pair compositions and/or functional properties (coding versus non-coding sequences, sequences that prefer to wrap around histones versus those that do not, etc.).

A better interpretation of fiber diffraction is particularly useful for understanding DNA–DNA interactions. Over 25 years ago, it was suggested how diffraction peaks determined by the interaxial spacing in fibers might be used for direct measurements of forces between DNA double helices (14). The present study shows how other peaks in the same fiber diffraction pattern may be used to reveal effects of these interactions on the double helix structure and, hence, illustrate the role of DNA structure in the interactions.

Together, advances in the diffraction and interaction theories provide important insights into the structure–interaction relationship for DNA, consistently explaining a large variety of experimental observations, as discussed below.

### Strand-groove alignment and DNA packing in cells and viruses

Analysis of the dependence of DNA fiber diffraction patterns on the spacing between neighboring molecules in hydrated fibers revealed a zipper-like alignment of strands and grooves on opposing double helices (25). Recently, such an alignment was visualized by cryoelectron microscopy of DNA aggregates inside viral particles (26). These observations demonstrate that the double helix structure is important for DNA–DNA interactions and DNA packing within cells and viruses.

Upon preferential binding of enough counterions in the major groove, electrostatic zipper attraction resulting from the alignment of the negatively charged strands with positively charged grooves (Figure 1C) may become sufficiently strong to cause DNA aggregation (28). Counterion-induced DNA aggregation has been observed in many studies and it appears essential for condensing DNA inside cells and viruses (16,35). Many models for aggregation have been proposed, but only the electrostatic zipper model involves strand-groove alignment (16). In our opinion, observations of such an alignment inside viral capsids (26) and good agreement between predicted (21) and measured (36–38) forces in DNA aggregates, as well as the counterion specificity of DNA aggregation (16,39), strongly support the zipper mechanism. On a cautionary note, we cannot discount some other mechanisms that might contribute to the overall effect (16).

### Helical coherence and structural adaptation of DNA

The strand-groove alignment is inhibited by sequence-related and thermal distortions in the helical coherence of DNA, yet the structural flexibility (plasticity) enables the double helices to adapt their structure to more favorable interactions for which they remain aligned ('Materials and Methods' section and Figure 2) (21,31). The present study shows how the helical coherence and structural adaptation of *B*-DNA might be evaluated from X-ray fiber diffraction patterns. The width of the diffraction peak of the fifth layer line (Figures 4 and 5) is not affected by the adaptation, but is solely determined by the helical coherence length $\lambda_c$ and the amplitude of DNA undulations [Equations (7) and (8) and Figure 5]. The

fitting of the dependence of the fifth layer line width on fiber hydration with Equations (7) and (8) supports this interpretation and yields a value of $\lambda_c$ that is in excellent agreement with estimates (15) based on NMR solution structures of short oligonucleotides (Figure 5, inset).

Information about the structural adaptation of *B*-DNA upon changes in fiber hydration might be extracted from the shapes of the diffraction peaks centered at the first two layer lines (Figure 4). However, this is a challenging task; requiring deconvolution of the corresponding molecular structure factors from intermolecular scattering [Equation (2)] and corrections for smearing of the diffraction peaks caused by imperfect ordering and orientation of DNA fibers. But even if we cannot accurately determine the characteristic adaptation length $\lambda_h^*$, it might be possible to study sharpening of the lower order layer lines qualitatively, which provides a signature of the onset of adaptation.

In the present study, we focused on *B*-DNA molecules, but a similar approach may be used for analysis of helical coherence and structural adaptation in other forms of DNA or RNA. Indeed, Equations (6–8) may be used for any type of double-helical nucleic acid, provided that the validity criteria for these approximations are satisfied, as discussed at the end of Appendix 1. Alternatively, the more general Equation (A5) might be used. For *B*-DNA, we selected the approximations based on previous estimates for the possible range of $\lambda_c$ (15) and $\lambda_h^*$ (21). Subsequent analysis of X-ray diffraction patterns validated these estimates and provided a more accurate value for $\lambda_c$ (Figure 5). The same approach may be used for X-ray diffraction studies of different kinds of *B*-DNA molecules, e.g. to study how $\lambda_c$ depends on the base pair composition and sequence. For other forms of DNA or for double helical RNA, the parameters that determine the validity of different approximations are not known *a priori*. However, the diffraction data might be compared with Equations (6–8 and A5) empirically, based on the observed shape of $k_z$-cross-sections of the diffraction maxima. The selected approximation might then be verified by extracting the relevant parameters from this comparison and substituting these parameters into the corresponding validity criteria.

### Pairing of homologous double helices

Proposed homology-based recognition and the pairing of intact, double-helical DNA molecules is a particularly important example of a potential biological role of the helical coherence and structural adaptation of *B*-DNA. The structural adaptation cost is lower for *B*-DNA double helices with homologous sequences, in which sequence-dependent variations in the base pair twist and rise are the same on opposing molecules and do not interfere with the alignment. Close juxtaposition should therefore be more energetically favorable for *B*-DNA molecules with identical or homologous sequences than for molecules with uncorrelated sequences (30). Preferential interactions between homologous double helices were directly observed in liquid crystalline aggregates (40), although additional experimental studies are needed to confirm this molecular mechanism for the observed homologue segregation.

Due to the lower adaptation cost, electrostatic zipper attraction may cause spontaneous pairing of homologous double helices in solution under conditions at which pairing of uncorrelated sequences is still unfavorable. Such pairing might contribute to an anomalous electrophoretic migration of homologous DNA fragments reported in Ref. (41). Furthermore, homologue pairing was recently directly observed in single molecule studies of *B*-DNA (42). [Note that some of the puzzling pairing features reported in Ref. (42) can be explained by the winding of unconstrained *B*-DNA molecules around each other into a braid (Cortini,R., Kornyshev,A.A., Lee,D.J. and Leikin,S., submitted for publication).]

The innate tendency of double-stranded *B*-DNA helices to form homologous pairs may play an important role in genetic recombination. Several recent studies suggested that some pairing of intact double helices may be involved in a homology search preceding genetic recombination in cells (43–46). The homology length requirement (>100 bp) predicted for the electrostatic zipper pairing (16,30) matches the homology length requirement for the recombination (47–50). One may speculate that transient electrostatic zipper pairing contributes to a 'coarse-grained' homology search (16,30,51), which is independent of double-strand breaks. Double helix invasion by a complementary single-stranded nucleoprotein filament may be responsible for a more precise recognition between shorter [10–50 bp (52,53)] sequences at a later stage in the recombination process, after the double-strand break. Of course, the very complex organization of the genome, particularly in eukaryotic cells, and the involvement of many different proteins may dramatically alter the pairing of homologous double helices. Nonetheless, nature often utilizes simple physics as a basis for very complex biological processes.

## CONCLUSIONS

- X-ray diffraction patterns from hydrated fibers and columnar aggregates of DNA contain information not only about the average structure of the double helix, but also about sequence-related and thermal fluctuations in base pair stacking, mutual alignment of the molecules and even intermolecular interactions.
- The more general fiber diffraction theory developed in the present work, which takes into account these features, may be utilized to study helical coherence and structural adaptation of nucleic acids.
- Analysis of *B*-DNA structure, packing and interactions based on this new, comprehensive theory of X-ray diffraction patterns leads to a more consistent picture of DNA physics and offers important clues for interpreting such phenomena as in-register DNA packing in viruses and homologous DNA pairing.
- This approach to X-ray diffraction could be extended to other long helical macromolecules and nanomaterials based on fibrous aggregates.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

This work was a part of a long project, for which we gratefully acknowledge many useful discussions with Steve Zimmerman and the use of the data taken from our previous joint paper.

## FUNDING

Leverhulme Trust grant F/07568, EPSRC grants EP/H010106/1 and EP/H004319/1, HFSP grant RGP0049/2010-C102 (to A.A.K.); Max-Planck Society (to D.J.L.); Intramural Research Program of NICHD (to S.L.); Alexander von Humboldt Foundation (to A.W.). Funding for open access charge: Intramural Research Program of NICHD, National Institutes of Health.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Franklin,R.E. and Gosling,R.G. (1953) Molecular configuration in sodium thymonucleate. *Nature*, **171**, 740–741.
2. Wilkins,M.H., Stokes,A.R. and Wilson,H.R. (1953) Molecular structure of deoxypentose nucleic acids. *Nature*, **171**, 738–740.
3. Watson,J.D. and Crick,F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
4. Cochran,W., Crick,F.H.C. and Vand,V. (1952) The structure of synthetic polypeptides. I. The transform of atoms on a helix. *Acta Cryst*., **5**, 581–586.
5. Klug,A., Crick,F.H.C. and Wyckoff,H.W. (1958) Diffraction by helical structures. *Acta Cryst*., **11**, 199–213.
6. Dickerson,R.E. and Drew,H.R. (1981) Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J. Mol. Biol*., **149**, 761–786.
7. Dickerson,R.E. (1992) DNA structure from A to Z. *Methods Enzymol*., **211**, 67–111.
8. Gorin,A.A., Zhurkin,V.B. and Olson,W.K. (1995) B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol*., **247**, 34–48.
9. Rozenberg,H., Rabinovich,D., Frolow,F., Hegde,R.S. and Shakked,Z. (1998) Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proc. Natl Acad. Sci. USA*, **95**, 15194–15199.
10. Olson,W.K. and Zhurkin,V.B. (2000) Modeling DNA deformations. *Curr. Opin. Struct. Biol*., **10**, 286–297.
11. Rhodes,D. and Klug,A. (1980) Helical periodicity of DNA determined by enzyme digestion. *Nature*, **286**, 573–578.
12. Wang,J.C. (1979) Helical repeat of DNA in solution. *Proc. Natl Acad. Sci. USA*, **76**, 200–203.
13. Zimmerman,S.B. and Pheiffer,B.H. (1979) Helical parameters of DNA do not change when DNA fibers are wetted: X-ray diffraction study. *Proc. Natl Acad. Sci. USA*, **76**, 2703–2707.
14. Rau,D.C., Lee,B. and Parsegian,V.A. (1984) Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel DNA double helices. *Proc. Natl Acad. Sci. USA*, **81**, 2621–2625.
15. Wynveen,A., Lee,D.J., Kornyshev,A.A. and Leikin,S. (2008) Helical coherence of DNA in crystals and solution. *Nucleic Acids Res*., **36**, 5540–5551.
16. Kornyshev,A.A., Lee,D.J., Leikin,S. and Wynveen,A. (2007) Structure and interactions of biological helices. *Rev. Mod. Phys*., **79**, 943–996.
17. Kornyshev,A.A. (2010) Physics of DNA: unravelling hidden abilities encoded in the structure of 'the most important molecule'. *Phys. Chem. Chem. Phys*., **12**, 12352–12378.
18. Barakat,R. (1987) X-ray scattering from helical structures possessing random variable twist. *Acta Cryst. A*, **43**, 45–49.
19. Stokes,D.L. and DeRosier,D.J. (1987) The variable twist of actin and its modulation by actin-binding proteins. *J. Cell Biol*., **104**, 1005–1017.
20. Inouye,H. (1994) X-ray scattering from a discrete helix with cumulative angular and translational disorders. *Acta Cryst. A*, **50**, 644–646.
21. Lee,D.J., Wynveen,A., Kornyshev,A.A. and Leikin,S. (2010) Undulations enhance the effect of helical structure on DNA interactions. *J. Phys. Chem. B*, **114**, 11668–11680.
22. Lucas,A.A. and Lambin,P. (2005) Diffraction by DNA, carbon nanotubes and other helical nanostructures. *Rep. Prog. Phys*., **68**, 1181–1249.
23. Zimmerman,S.B. and Pheiffer,B.H. (1983) DNA structures and transitions: X-ray diffraction studies of solvated fibers. *Cold Spring Harb. Symp. Quant. Biol*., **47**, 67–76.
24. Langridge,R., Wilson,H.R., Hooper,C.W., Wilkins,M.H.F. and Hamilton,L.D. (1960) Molecular configuration of deoxyribonucleic acid. 1. X-ray diffraction study of a crystalline form of the lithium salt. *J. Mol. Biol*., **2**, 19–37.
25. Kornyshev,A.A., Lee,D.J., Leikin,S., Wynveen,A. and Zimmerman,S.B. (2005) Direct observation of azimuthal correlations between DNA in hydrated aggregates. *Phys. Rev. Lett*., **95**, 148102.
26. Leforestier,A. and Livolant,F. (2009) Structure of toroidal DNA collapsed inside the phage capsid. *Proc. Natl Acad. Sci. USA*, **106**, 9157–9162.
27. Kornyshev,A.A. and Leikin,S. (1997) Theory of interaction between helical molecules. *J. Chem. Phys*., **107**, 3656–3674.
28. Kornyshev,A.A. and Leikin,S. (1999) Electrostatic zipper motif for DNA aggregation. *Phys. Rev. Lett*., **82**, 4138–4141.
29. Hud,N. (ed.), (2009) *Nucleic Acid-metal Ion Interactions*. RSC Publishing, Cambridge.
30. Kornyshev,A.A. and Leikin,S. (2001) Sequence recognition in the pairing of DNA duplexes. *Phys. Rev. Lett*., **86**, 3666–3669.
31. Cherstvy,A.G., Kornyshev,A.A. and Leikin,S. (2004) Torsional deformation of double helix in interaction and aggregation of DNA. *J. Phys. Chem. B*, **108**, 6508–6518.
32. Griffith,J.D. (1978) DNA structure: evidence from electron microscopy. *Science*, **201**, 525–527.
33. Levitt,M. (1978) How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proc. Natl Acad. Sci. USA*, **75**, 640–644.
34. Vainshtein,B.K. (1966) *Diffraction of X-rays by Chain Molecules*. Elsevier, Amsterdam.
35. Bloomfield,V.A. (1997) DNA condensation by multivalent cations. *Biopolymers*, **44**, 269–282.
36. Rau,D.C. and Parsegian,V.A. (1992) Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces. *Biophys. J*., **61**, 246–259.
37. Podgornik,R., Strey,H.H., Gawrisch,K., Rau,D.C., Rupprecht,A. and Parsegian,V.A. (1996) Bond orientational order, molecular motion, and free energy of high-density DNA mesophases. *Proc. Natl Acad. Sci. USA*, **93**, 4261–4266.
38. Todd,B.A., Parsegian,V.A., Shirahata,A., Thomas,T.J. and Rau,D.C. (2008) Attractive forces between cation condensed DNA double helices. *Biophys. J*., **94**, 4775–4782.
39. DeRouchey,J., Parsegian,V.A. and Rau,D.C. (2010) Cation charge dependence of the forces driving DNA assembly. *Biophys. J*., **99**, 2608–2615.
40. Baldwin,G.S., Brooks,N.J., Robson,R.E., Wynveen,A., Goldar,A., Leikin,S., Seddon,J.M. and Kornyshev,A.A. (2008) DNA double helices recognize mutual sequence homology in a protein free environment. *J. Phys. Chem. B*, **112**, 1060–1064.
41. Inoue,S., Sugiyama,S., Travers,A.A. and Ohyama,T. (2007) Self-assembly of double-stranded DNA molecules at nanomolar concentrations. *Biochemistry*, **46**, 164–171.
42. Danilowicz,C., Lee,C.H., Kim,K., Hatch,K., Coljee,V.W., Kleckner,N. and Prentiss,M. (2009) Single molecule detection of

direct, homologous, DNA/DNA pairing. *Proc. Natl Acad. Sci. USA*, **106**, 19824–19829.
43. Aragon-Alcaide,L. and Strunnikov,A.V. (2000) Functional dissection of in vivo interchromosome association in Saccharomyces cerevisiae. *Nat. Cell Biol.*, **2**, 812–818.
44. Barzel,A. and Kupiec,M. (2008) Finding a match: how do homologous sequences get together for recombination? *Nat .Rev. Genet.*, **9**, 27–37.
45. Molnar,M. and Kleckner,N. (2008) Examination of interchromosomal interactions in vegetatively growing diploid Schizosaccharomyces pombe cells by Cre/loxP site-specific recombination. *Genetics*, **178**, 99–112.
46. Zickler,D. and Kleckner,N. (1999) Meiotic chromosomes: integrating structure and function. *Annu. Rev. Genet.*, **33**, 603–754.
47. Singer,B.S., Gold,L., Gauss,P. and Doherty,D.H. (1982) Determination of the amount of homology required for recombination in bacteriophage T4. *Cell*, **31**, 25–33.
48. Rubnitz,J. and Subramani,S. (1984) The minimum amount of homology required for homologous recombination in mammalian cells. *Mol. Cell Biol.*, **4**, 2253–2258.
49. Watt,V.M., Ingles,C.J., Urdea,M.S. and Rutter,W.J. (1985) Homology requirements for recombination in Escherichia coli. *Proc. Natl Acad. Sci. USA*, **82**, 4768–4772.
50. Shen,P. and Huang,H.V. (1986) Homologous recombination in Escherichia coli: dependence on substrate length and homology. *Genetics*, **112**, 441–457.
51. Kornyshev,A.A. and Wynveen,A. (2009) The homology recognition well as an innate property of DNA structure. *Proc. Natl Acad. Sci. USA*, **106**, 4683–4688.
52. Hsieh,P., Camerini-Otero,C.S. and Camerini-Otero,R.D. (1992) The synapsis event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA. *Proc. Natl Acad. Sci. USA*, **89**, 6492–6496.
53. Rothenberg,E., Grimme,J.M., Spies,M. and Ha,T. (2008) Human Rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes. *Proc. Natl Acad. Sci. USA*, **105**, 20274–20279.
54. Lee,D.J., Leikin,S. and Wynveen,A. (2010) Fluctuations and interactions of semi-flexible polyelectrolytes in columnar assemblies. *J. Phys. Condens. Matter*, **22**, 72202.
55. Odijk,T. (1983) On the statistics and dynamics of confined or entangled stiff polymers. *Macromolecules*, **16**, 1340–1344.
56. Odijk,T. (1986) Theory of lyotropic polymer liquid-crystals. *Macromolecules*, **19**, 2314–2329.
57. Bolshoy,A., McNamara,P., Harrington,R.E. and Trifonov,E.N. (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl Acad. Sci. USA*, **88**, 2312–2316.
58. Marko,J.F. and Siggia,E.D. (1995) Stretching DNA. *Macromolecules*, **28**, 8759–8770.

## APPENDIX 1: THEORY OF MOLECULAR STRUCTURE FACTORS IN NON-CRYSTALLINE COLUMNAR ASSEMBLIES OF DNA

### Molecular structure factors

The molecular structure factor is defined as

$$s(k_z, K, n) = \left\langle \upsilon(k_z, K, n)\upsilon(-k_z, K, -n)\right\rangle \tag{A1}$$

where

$$\upsilon(k_z, K, n) = \frac{1}{2\pi J_n(Ka)}\int_0^{2\pi} d\phi \int_{-\infty}^{\infty} dz \int_0^{\infty} r dr\, J_n(Kr) n_p(r, \phi, z) e^{in\phi} e^{ik_z z} \tag{A2}$$

and $n_p(r, \phi, z)$ is the density of phosphate groups in cylindrical coordinates with the $z$-axis being along the central axis of the molecule.

To calculate $s(k_z, K, n)$ in DNA fibers, we approximate the phosphate density by

$$n_p(\mathbf{R}, z) = \int d\mathbf{R}'\, \delta(\mathbf{R} - \mathbf{R}' - \mathbf{r}(z))\, n_p^{(0)}(\mathbf{R}', z); \tag{A3}$$

where $\mathbf{r}(z)$ is the lateral displacement of the DNA centerline from the vertical axis (Figure 1A)

$$n_p^{(0)}(r, \phi, z) = \frac{1}{a}\delta(r - a) \times \sum_j \left\{\delta\left(\phi - \tilde{\phi}_s - \phi_j\right)\delta\left(z - z_j\right) + \delta\left(\phi + \tilde{\phi}_s - \phi_j\right)\delta\left(z - z_j\right)\right\} \tag{A4}$$

is the phosphate density for a straight DNA molecule; $\phi_j = \phi_0 + \sum_{i=1}^{j} \Omega_i$ and $z_j = z_0 + \sum_{i=1}^{j} h_i$ are the azimuthal orientation and axial coordinate of the base pair $j$, respectively; $\Omega_i$ and $h_i$ are the twist angle and axial rise between the adjoining base pairs $i-1$ and $i$, respectively; $\delta(x)$ is the Dirac delta-function. For simplicity, we explicitly account for lateral displacements (bending) of the DNA centerline and for variations in the base pair twist and rise relative to the centerline, but neglect intra-base pair distortions (such as propeller twist). We assume the lateral displacements to be small and independent of the twist and rise. For a justification of these approximations, see Supplementary Data.

Substitution of Equations (A3) and (A4) into Equations (A1) and (A2) and calculation of the averages yields

$$s(k_z, K, n) = \frac{\left[\cos n\tilde{\phi}_s\right]^2}{\pi^2}\sum_{i,j}\cos\left(j\bar{h}(k_z + n\bar{g})\right) \times \exp\left(-\frac{n^2}{2}\left\langle\left(\Phi_i - \Phi_{i+j}\right)^2\right\rangle - \frac{K^2}{4}\left\langle\left(\mathbf{r}_i - \mathbf{r}_{i+j}\right)^2\right\rangle\right). \tag{A5}$$

Here the summation is performed over all base pairs in the x-ray beam and $\mathbf{r}_i$ is the lateral displacement of the centerline at base pair $i$.

$$\Phi_i \equiv \Phi_0 + \sum_{m=1}^{i}(\Omega_m - \bar{g}h_m) \tag{A6}$$

is the helical phase at base pair $i$ [for a derivation of Equation (A5), see Supplementary Data]. The correlation functions for lateral displacements, $\left\langle\left(\mathbf{r}_i - \mathbf{r}_{i+j}\right)^2\right\rangle$ and helical phase variations, $\left\langle\left(\Phi_i - \Phi_{i+j}\right)^2\right\rangle$, are discussed below. The average value of the helical phase $<\Phi_i> = \Phi_0 \equiv \phi_0$ characterizes the azimuthal orientation of the whole molecule. Variations in $\Phi_i$ characterize the effects of sequence-related variations and thermal fluctuations in $\Omega_m$ and $h_m$ on the local periodicity of the double helix (15). Equation (A5) reduces the problem to a calculation of the pair correlation functions for the lateral displacement and helical phase, which were recently described in Ref. (21).

### Undulations

The pair correlation function for the lateral displacement is given by (54)

$$\left\langle \left(\mathbf{r}_i - \mathbf{r}_{i+j}\right)^2 \right\rangle \approx 2d^2\Psi\left(|j|\bar{h}/\lambda_B\sqrt{2}\right), \quad \textbf{(A7)}$$

where $d$ is the root mean square amplitude of thermal undulations

$$\Psi(t) \equiv 1 - \sqrt{2}\exp(-t)\cos\left(t - \frac{\pi}{4}\right), \quad \textbf{(A8)}$$

and the so-called deflection length,

$$\lambda_B \approx \left(\sqrt{2}d^2 l_p^b\right)^{1/3}, \quad \textbf{(A9)}$$

is the characteristic length along the $z$-axis after which $\mathbf{r}_i$ can no longer accumulate in one direction due to interactions with neighbors and the molecule's centerline is deflected back toward its average position (54–56). Here we used the fact that the DNA bending persistence length for thermal undulations ($l_p^b \approx 500$ Å) is much shorter than the corresponding length for sequence-related bending (57,58) and, therefore, neglected the contribution of the latter.

### Helical phase fluctuations

The pair correlation function for helical phase variations is complicated by the structural adaptation effects, as schematically illustrated in Figure 2 (21). However, this correlation function may be modeled based on the following arguments. At $|z_{i+j} - z_i| << \lambda_h^*$, this function is not affected by the structural adaptation and $\left\langle \left(\Phi_i - \Phi_{i+j}\right)^2 \right\rangle \approx |j|\bar{h}/\lambda_c$ (15,31). At $|z_{i+j} - z_i| >> \lambda_h^*$, realignment of the neighboring molecules through the structural adaptation prevents the accumulation of helical phase variations with increasing $|j|$. For infinitely long double helices without structural defects, the correlation function predicted by our calculations levels off to $\left\langle \left(\Phi_i - \Phi_{i+j}\right)^2 \right\rangle \approx \lambda_h^*/\lambda_c$ (21,31). In other words, realignment of molecules through the structural adaptation suppresses long-range fluctuations in the helical phase (Figure 2C, lower left). In contrast, disruptions in the sugar-phosphate backbone and other essential structural defects realign the molecules by introducing disruptions in the helical phase (Figure 2C, lower right). Provided that these defects have a normal distribution and may be described by Gaussian statistics, we may then expect $\left\langle \left(\Phi_i - \Phi_{i+j}\right)^2 \right\rangle \approx |j|\bar{h}/\lambda_d$ at very large $|z_{i+j} - z_i|$. While this random-walk correlation function is similar to the one at $|z_{i+j} - z_i| << \lambda_h^*$, the corresponding correlation length $\lambda_d$ is determined by the distribution of disruptive structural defects in the double helix. We expect $\lambda_d$ to be of the order of the average length between such defects ($\sim 10^3$–$10^4$ bp), i.e. $\lambda_d >> \lambda_c$. These features of the helical phase correlation function are captured by the following interpolation formula

$$\left\langle \left(\Phi_i - \Phi_{i+j}\right)^2 \right\rangle \approx |j|\frac{\bar{h}}{\lambda_c}\theta\left(\lambda_h^* - |j|\bar{h}\right) + \left(\frac{\lambda_h^*}{\lambda_c} + \frac{|j|\bar{h} - \lambda_h^*}{\lambda_d}\right)\theta\left(|j|\bar{h} - \lambda_h^*\right), \quad \textbf{(A10)}$$

which yields the expected behavior of the correlation function at $|j|\bar{h} << \lambda_h^*, \lambda_h^* << |j|\bar{h} << \lambda_d$ and $|j|\bar{h} >> \lambda_d$. Here $\theta(x)$ is the Heaviside step function [$\theta(x) = 0$ at $x < 0$ and $\theta(x) = 1$ at $x > 0$].

### Simplifying approximations

Equations (3), (A5), (A7) and (A10) fully define the model for calculating molecular scattering intensities for non-ideal, undulating DNA double helices. Generally the summation in Equation (A5) has to be performed numerically, but analytical approximations may be obtained at $n^2K^2d^2 << 2\sqrt{2}\lambda_c/\lambda_B$ and at $2\lambda_c/\lambda_h^* << n^2 << 2\lambda_c/h$ (Supplementary Data). In the first case, we may use $\left\langle \left(\mathbf{r}_i - \mathbf{r}_{i+j}\right)^2 \right\rangle \approx 2d^2$ and the summation in Equation (A5) yields Equation (6). Note that as $d \to 0, \lambda_B \to 0$, Equation (6) becomes valid at all $n$. In the second case, the summation yields Equations (7) and (8) (Supplementary Data), where

$$w_0(K,n,d) = \sqrt{\frac{\left(\sum_{j=-\infty}^{\infty}\exp\left(-|j|\frac{n^2h}{2\lambda_c}\right)\exp\left[-\frac{K^2d^2}{2}\Psi\left(|j|\frac{\bar{h}}{\lambda_B\sqrt{2}}\right)\right]\right)^3}{2\sum_{j=-\infty}^{\infty}j^2\exp\left(-|j|\frac{n^2h}{2\lambda_c}\right)\exp\left[-\frac{K^2d^2}{2}\Psi\left(|j|\frac{\bar{h}}{\lambda_B\sqrt{2}}\right)\right]}} \quad \textbf{(A11)}$$

and

$$w_1(K,n,d) = \sqrt{\frac{\sum_{j=-\infty}^{\infty}j^2\frac{n^4\bar{h}^2}{4\lambda_c^2}\exp\left(-|j|\frac{n^2h}{2\lambda_c}\right)\exp\left[-\frac{K^2d^2}{2}\Psi\left(|j|\frac{\bar{h}}{\lambda_B\sqrt{2}}\right)\right]}{2\sum_{j=-\infty}^{\infty}\exp\left(-|j|\frac{n^2h}{2\lambda_c}\right)\exp\left[-\frac{K^2d^2}{2}\Psi\left(|j|\frac{\bar{h}}{\lambda_B\sqrt{2}}\right)\right]}}. \quad \textbf{(A12)}$$

In the case of $B$-DNA, we expect Equations (6–8) to cover all cases of practical importance. Indeed, previous estimates suggest that $\lambda_c \sim 100$ Å and $2\lambda_c/h \sim 60$ (15). High-quality diffraction on $B$–DNA fibers is usually observed at interaxial distances $R_0$ between neighboring molecules from 24 to 40 Å. Calculations of the values of $d$ and $\lambda_h^*$ for $B$-DNA, based on the model described in (21), suggest that at physiological salt concentrations (0.1–0.2 M NaCl) Equation (6) should be a good approximation at $n = \pm 1, \pm 2$ and Equations (7) and (8) should be a good approximation at $n = \pm 5$. For $n = \pm 3$, we expect Equation (6) to work at $R_0$ smaller than $\approx 28$ Å and Equations (7) and (8) at larger $R_0$. Both approximations yield closely matching results for $s(k_z, K, \pm 3)$ at $R_0 \approx 28$ Å.