PAPER • OPEN ACCESS

Thermal denaturation of A-DNA

To cite this article: J Valle-Orero *et al* 2014 *New J. Phys.* **16** 113017

New Journal of Physics
The open access journal at the forefront of physics
Deutsche Physikalische Gesellschaft DPG | IOP Institute of Physics

Thermal denaturation of A-DNA

J Valle-Orero[1,2,3], A R Wildes[1], N Theodorakopoulos[4,5], S Cuesta-López[2,6], J-L Garden[7], S Danilkin[8] and M Peyrard[2]

[1] Institut Laue-Langevin, CS 20156, 71 avenue des Martyrs, F-38042 Grenoble Cedex 9, France
[2] Ecole Normale Supérieure de Lyon, Laboratoire de Physique CNRS UMR 5672, 46 allée d'Italie, F-69364 Lyon Cedex 7, France
[3] Department of Biological Sciences, Columbia University, New York, NY 10027, USA
[4] Theoretical and Physical Chemistry Institute, National Hellenic Research Foundation, Vasileos Constantinou 48, 116 35 Athens, Greece
[5] Fachbereich Physik, Universität Konstanz, D-78457 Konstanz, Germany
[6] University of Burgos, Hospital del Rey s/n, S-09001, Burgos, Spain
[7] CNRS et Université Grenoble Alpes, Institut NEEL, F-38042 Grenoble, France
[8] ANSTO, Locked Bag 2001, Kirrawee DC, NSW 2232, Australia
E-mail: wildes@ill.fr

Received 27 May 2014, revised 26 August 2014
Accepted for publication 23 September 2014
Published 7 November 2014
New Journal of Physics **16** (2014) 113017
doi:10.1088/1367-2630/16/11/113017

Abstract

The DNA molecule can take various conformational forms. Investigations focus mainly on the so-called 'B-form', schematically drawn in the famous paper by Watson and Crick [1]. This is the usual form of DNA in a biological environment and is the only form that is stable in an aqueous environment. Other forms, however, can teach us much about DNA. They have the same nucleotide base pairs for 'building blocks' as B-DNA, but with different relative positions, and studying these forms gives insight into the interactions between elements under conditions far from equilibrium in the B-form. Studying the thermal denaturation is particularly interesting because it provides a direct probe of those interactions which control the growth of the fluctuations when the 'melting' temperature is approached. Here we report such a study on the 'A-form' using calorimetry and neutron scattering. We show that it can be carried further than a similar study on B-DNA, requiring the improvement of thermodynamic models for DNA.

Keywords: DNA melting, statistical mechanics, neutron scattering, nonlinear lattice dynamics

1. Introduction

The A-form of DNA played a major role in the early attempts to understand DNA structure. It was actually an x-ray diagram from A-DNA, shown at a conference in 1951 by M Wilkins, that gave Watson and Crick [1] the key clue to solving the structure of the B-form [2]. Franklin and Gosling worked extensively on A-DNA [3] and with good reason: DNA fibers, which are very convenient to obtain oriented DNA samples, crystallize better and give rise to higher quality x-ray diffraction patterns when in the A-form.

DNA fibers are still the object of structural studies nowadays, however most investigations on DNA are performed in solution. This presents certain limits to the study of the thermodynamics of the molecule. DNA adopts only the B-form in an aqueous solution, and studies of the thermal stability will only provide the fraction of open base pairs as a function of temperature. Such studies have no access to the spatial structure of the DNA molecules. Therefore they cannot determine how the open and closed regions are distributed.

Thermal measurements on A-DNA are important to test the influence of the conformational form on the melting transition and the universality of models used to describe DNA thermodynamics. DNA adopting the A-form may also be an important part of the gene transcription process *in vitro* [4]. It is possible to have A-form in solution with high alcohol concentrations [5]. However, the DNA often takes a combination of A and B-forms, as well as molecules with B–A interfaces, so that experimental results cannot be quantitatively interpreted. A-DNA can be prepared from randomly oriented fibers and studies of the melting process have been reported [6]. Neither DNA in solution nor in randomly oriented fibers will give unambiguous experimental information on the spatial structure.

We recently used neutron scattering from highly oriented B-DNA fibers [7, 8] to show that structural information on the thermal denaturation transition can be obtained. The position of the Bragg peaks associated to base-pair stacking gives access to the molecular structure and the width provides a measure of the size of the closed regions. This is crucial information for building a theory of the DNA melting phase transition. Various models can lead to the same melting curve but, if spatial information on the size of the closed clusters versus temperature is included, the experiments can discriminate between models. We were able to follow the early part of the melting transition in B-DNA, however the fiber structure collapses before the full denaturation of the molecules preventing the collection of structural data through the entire transition.

Here we show that the same type of study can be performed for A-DNA fibers that are free of B-DNA contamination. Scattering data show the structure, thus our experiments prove that DNA is indeed in pure A-form. The structure of the A-DNA samples does not collapse and thus the study can be carried further through the transition than with B-form. The spatial information provides a detailed view of the melting transition up to its last stage, yielding data that required further development of DNA models for a quantitative analysis.

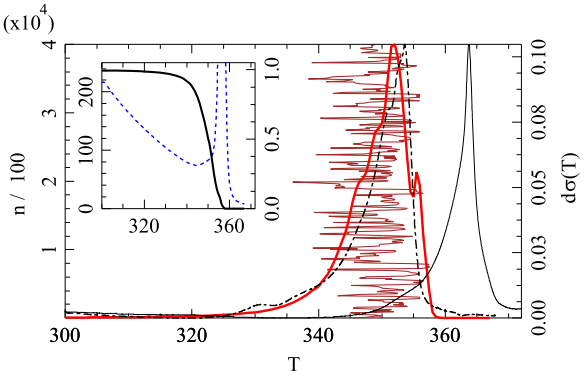

Figure 1. Comparison between the melting profile of A-DNA films obtained by calorimetry (dashed–dotted line) and the theoretical profile given by fitting with the modified PBD model (thick red line) for a sequence of 40 000 base pairs extracted from a salmon sequence listed in the GenBank [23]. For comparison the thin (black) full line shows the specific heat of a B-DNA film, made of Li–DNA humidified under 92% relative humidity. Both calorimetric curves have been normalized to have the same maximum as the theoretical melting profile to allow a better comparison of the shapes of the curves. The zig-zag brown line shows the melting map of the sequence analyzed theoretically. The vertical axis extends along the 40 000 base pairs (left scale). The sequence is divided in chunks of 100 base pairs and the horizontal position of a point along the line gives the opening temperature of each chunk. The inset shows the closed fraction (full black line, right scale) and the average size of the closed clusters (blue dash line, left scale) versus temperature.

2. Experimental results

2.1. The sample

The samples were prepared using the 'wet spinning' technique [8, 9] which produces films made of parallel fibers in which the molecules themselves are aligned. The films were prepared from a solution of salmon testes DNA with sodium counterions (1.58 g l^{-1} DNA in 0.15 M NaCl). Electrophoresis on a non-denaturing gel with a small piece of the sample dissolved in water showed the length of the DNA to be $\gtrsim$ 20 kB.

The films were dried and then humidified in a ^2H_2O atmosphere with 56% relative humidity. This hydration leads largely to the A-form for Na–DNA. A thermal treatment is necessary to suppress a small B-DNA contamination and to achieve a higher degree of order in the sample [13]. Diffraction patterns confirmed that the sample had a pure A-form.

In comparison, the samples used for our previous study were made from a lithium solution (0.6 g l^{-1} DNA in 0.4 M LiCl), with the water content of the samples being set in a 75% relative humidity atmosphere [7, 8]. These conditions gave fibre DNA in a pure B-form.

2.2. Calorimetric denaturation curve

Differential scanning calorimetry (DSC) studies were performed using a Seratam Micro DSC III calorimeter. Samples cut from A-form DNA fiber films were hermetically sealed in a Hastelloy sample tube, and the calorimetry measurements were made relative to an empty tube. Measurements were made for a variety of heating rates between 0.6 and 1.2 K min^{-1}, with the data showing no discernible difference between ramp rates.

Denaturation curves for A- and B-form fiber DNA, recorded by differential scanning calorimetry, are plotted in figure 1. Both curves show a clear peak associated with the melting transition for the films, albeit at different temperatures. The melting temperature is known to vary considerably as a function of the counterion species and concentration [10–12], and may also be influenced by the conformation. However, the difference in the value of the melting temperature is not significant for the subsequent analysis. More meaningful is the observation that, despite the two samples being prepared from salmon DNA with the same average sequence, the width of the A-DNA denaturation peak is significantly broader than that for the B-DNA.

2.3. Neutron diffraction results

The structure of oriented DNA as it denatures can be followed with neutron diffraction by recording the temperature evolution of selected Bragg peaks. To monitor base-pair openings the Bragg peaks must be particularly sensitive to the internal structure of the DNA molecule, rather than to the crystalline structure of the fibers.

Preliminary experiments were performed at the Bragg Institute (ANSTO) Sydney using the TAIPAN three-axis spectrometer. A second experiment was performed at the Institut Laue-Langevin, France, using the three-axis spectrometer IN8. The DNA films were sealed between aluminum plates using lead wire inside a humidity chamber to ensure the conservation of water molecules within the sample. The instruments were configured with a graphite monochromator with an initial wavevector $k_i = 2.662$ Å^{-1} for the low-Q peak and 4.1 Å^{-1} for the high-Q peaks. No energy analysis was used for the measurements. The temperature was raised from 295 K to 360 K using a heated nitrogen jet and the stability for the duration of each measurement was <0.1 K. The samples were therefore considered to be in thermal equilibrium at each temperature. We ensured that the configuration of the instruments did not change during the measurements by scanning the aluminium (111) Bragg peak at all temperatures.

The denaturation transition is irreversible. The sharp Bragg peaks do not reappear on cooling, meaning that we could not investigate how the melted DNA recombines.

The right half of figure 2 shows the neutron diffraction pattern from a sample of A-form DNA. It displays three intense peaks, characteristic of A-DNA [3, 14], at positions $Q_{||} = 6, 7, 8 \times 2\pi/d$, where $d = 2.54$ Å is the distance separating base pairs along the helix axis, and $Q_{||}$ is the reciprocal lattice vector parallel to this axis. A sharp Bragg peak is also present at $Q_{||} = 2 \times 2\pi/d$. The peaks are reproduced in a calculation of the neutron diffraction from a single A-DNA molecule, shown on the left half of figure 2, proving that they predominately probe the molecular structure rather than the crystallinity. These peaks were chosen to follow the A-DNA structure versus temperature. We henceforth define the set of three peaks at $Q_{||} = 6, 7, 8 \times 2\pi/d$ as 'high-Q' peaks and the peak at $Q_{||} = 2 \times 2\pi/d$ as 'low-Q'.

The peaks were measured in two scans parallel to $Q_{||}$. The evolution of the low-Q peak was followed in a scan centered at $\left(Q_{\perp}, Q_{||}\right) = (0.23, 0.5)$, while the three high-$Q$ peaks were followed in a scan centered at $\left(Q_{\perp}, Q_{||}\right) = (0.6, 1.8)$. Figure 3 shows representative scans at two temperatures.

The temperature evolution of the A- and B-DNA films show an important difference. The fiber structure of B-DNA films 'collapses' before the full denaturation, becoming liquid-like and losing its Bragg peaks. The collapse can be understood as an entropic effect associated with the high flexibility of denatured regions of the double helix, destabilizing the whole film structure well before the complete denaturation. The loss of the Bragg peaks meant that the

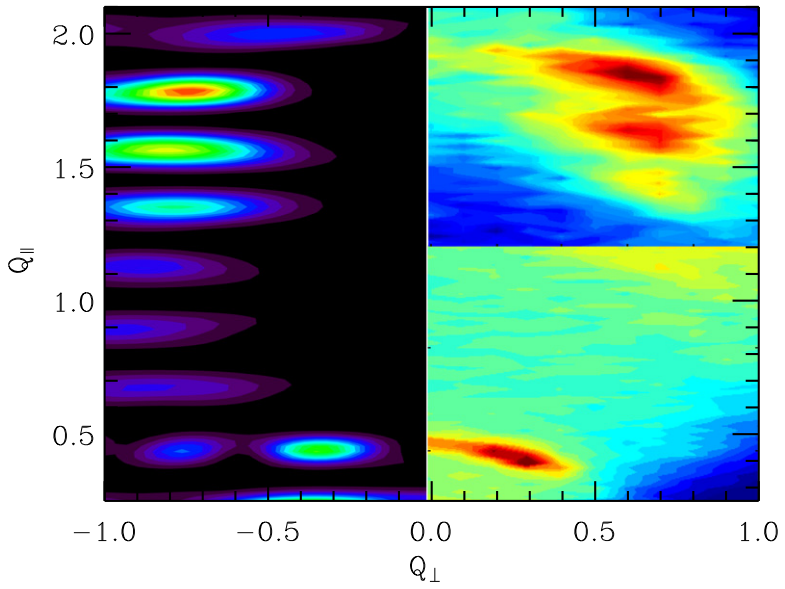

Figure 2. Comparison between the experimental neutron scattering intensities (right half) and theoretical structure factor (left half) for oriented A-DNA. The reciprocal space direction $Q_{\|}$ is parallel to the molecular axis and $Q_{\perp}$ is orthogonal to that axis. Their units are in Å^{-1}. The theory calculation was performed for two turns of an ideal A-DNA molecule, accounting for the neutron scattering length of each atom. The experimental data were obtained in two measurements with different wavelengths, optimized for the different Q domains, hence the absolute intensities cannot be directly compared. Nevertheless, the qualitative picture clearly shows that the experimental reciprocal space map is mostly determined by the form factor of a single molecule.

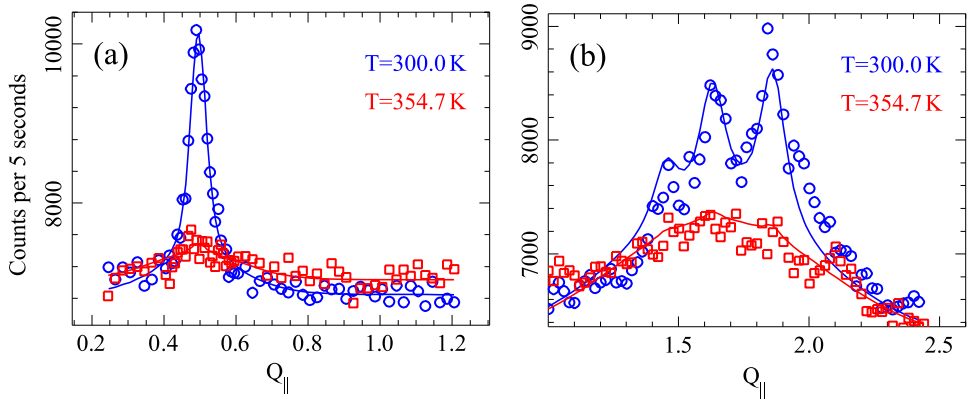

Figure 3. Examples of the neutron scattering intensity as a function of $Q_{\|}$ for (a) the low-Q and (b) high-Q Bragg peaks. The Q units are in Å^{-1}. The scans were made at constant $Q_{\perp} = 0.23$ Å^{-1} and $Q_{\perp} = 0.6$ Å^{-1} for the low- and high-Q peaks respectively. Examples of fits, as explained in the text, are also shown.

neutron scattering experiments could not follow the evolution through the complete transition [8]. After thermal denaturation and cooling, the B-DNA films exhibit glassy properties [15].

Fiber collapse during denaturation does not occur for our highly oriented A-DNA samples, as confirmed by optical microscopy. This may be because the higher diameter of the molecules leads to a higher rigidity, reducing the film's tendency to collapse by entropic

effects, and also because the crystalline organization of the molecules is more compact in A-DNA than in B-DNA as there is less interstitial water, leaving less space for large orientational fluctuations of molecular segments. The orientation of the fibers is not lost until thermal denaturation is complete, and we could follow the Bragg peaks through the complete transition.

It is likely that sample quality also plays a role. A recent study on non-oriented fibers showed very similar Bragg peak behaviour in both A- and B-DNA [6], with both data sets resembling our measurements on B-DNA [8]. It is feasible that the previous study also suffered from fiber collapse, hence explaining the similarity of the two data sets. The higher degree of order in our wet-spun samples, as well as the heat treatment that removed B-form contamination, not only gives unambiguous information on the A-DNA structure, it also appears to stabilize the fibers to higher temperature.

The Bragg peaks represent coherent scattering from closed regions of A-DNA, and their analysis gives quantitative estimates of the quantity and size of those closed regions. The peaks were fitted with Lorentzians. Careful attention needed to be paid to the background for the fits. In addition to the strong incoherent scattering caused by hydrogen in the sample, figure 3 shows that diffuse peaks persist in both Q ranges at high temperatures. Analysis shows that these features are present at all temperatures and are probably due to a liquid-like structure factor from amorphous regions of the sample. The humps may be regarded as background. They were fitted with Gaussian peaks at high temperature and then the parameters were fixed for all subsequent fits.

The low-Q peaks were fitted by the function

$$N_{1\mathrm{p}}(Q) = \frac{I_{\mathrm{low}}}{\pi}\frac{\Gamma_{\mathrm{low}}/2}{\left(\Gamma_{\mathrm{low}}/2\right)^2 + \left(Q - Q^{\mathrm{c}}_{\mathrm{low}}\right)^2} \tag{1}$$

$$+I' \exp\left[-\frac{\left(Q - Q^{\mathrm{c}}_{\mathrm{low}}\right)^2}{2\Gamma'^2}\right] + B'_{\mathrm{gnd}}, \tag{2}$$

while the set of three high-Q peaks were fitted with

$$N_{3\mathrm{p}}(Q) = \sum_{i=1,3}\frac{I_i}{\pi}\frac{\Gamma_{\mathrm{high}}/2}{\left(\Gamma_{\mathrm{high}}/2\right)^2 + \left(Q - Q^{\mathrm{c}}_i\right)^2} \tag{3}$$

$$+I'' \exp\left[-\frac{\left(Q - Q^{\mathrm{c}}_{\mathrm{high}}\right)^2}{2\Gamma''^2}\right] + B''_{\mathrm{gnd}}. \tag{4}$$

In these expressions I designates the integrated intensity of each peak, Γ its width in units of $\mathrm{Å}^{-1}$, and Q^{c} the position of its center. The experimental resolution was measured using a silicon Bragg peak. The measurements were used to parameterize the instrument resolution for all Q and the DNA data were fitted by model expressions convoluted with the expression for the resolution.

Examples of the fits are shown in figure 3. We applied certain constraints to the variables in order to have robust and systematic fits, which are particularly problematic when close to the melting temperature where the signal-to-noise is small. The widths of the three high-Q peaks

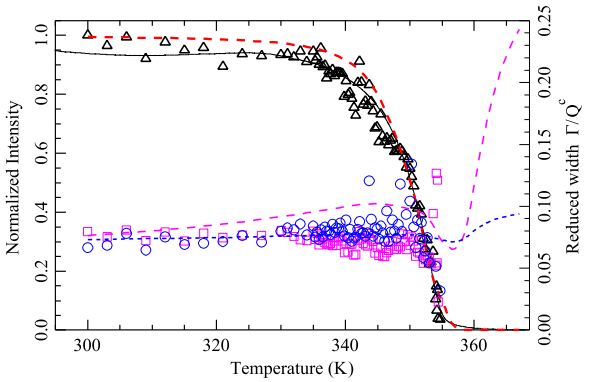

Figure 4. Neutron scattering data in comparison with calorimetric results and the theoretical model of the structure factor. The full (black) line shows the unmelted fraction of the DNA sample versus temperature as given by calorimetry. A linear variation obtained from the highest and lowest temperatures of the calorimetric curve, $C_p(T)$, was first subtracted from the data. The remaining peak was normalized to a unit area, and then we integrated the curve from the lowest temperature to T to get the fraction of open DNA. Subtracting this quantity from 1 gives the fraction of closed DNA. The triangular symbols show the intensity of the Bragg peaks normalized to $I_{\mathrm{low}}\,(300\ \mathrm{K})$. The dashed (red) line shows the melting curve fitted with the PBD model, derived from the data in figure 1. The open (pink) squares show the ratio of the width of the low-Q Bragg peaks to the peak center, $\Gamma_{\mathrm{low}}/Q^{\mathrm{c}}_{\mathrm{low}}$. The open (blue) circles show $\Gamma_{\mathrm{high}}/Q^{\mathrm{c}}_{\mathrm{high}}$, the width of the Bragg peaks at high-Q divided by the position of the middle peak. The dashed (pink) line and the dotted (blue) line show the theoretical values of $\Gamma_{\mathrm{low}}/Q^{\mathrm{c}}_{\mathrm{low}}$ and $\Gamma_{\mathrm{high}}/Q^{\mathrm{c}}_{\mathrm{high}}$ respectively.

were constrained to be equal, defined as Γ_{high}. We justify this by noting that the three peaks are close together in Q and that their widths, as confirmed by close inspection of the calculation in figure 2, are very similar. The data at 300 K were then fitted. The resulting values of Q^{c} were then fixed for all temperatures. The fitted values for I at 300 K were used to calculate the I/I_{low} ratios for the three high-Q peaks, where I_{low} is the integrated intensity of the low-Q peak, and these ratios were fixed for all subsequent fits. We justify this by noting that the Bragg peaks come from closed regions of DNA. The spatial correlation length may change, but the integrated intensities will be a result of the number of scattering centers in the closed regions. Their ratios will not vary substantially with temperature. The number of free parameters to fit the temperature-dependent data thus reduces to three: $I_{\mathrm{low}}(T)$; $\Gamma_{\mathrm{high}}(T)$; and the width of the low-Q peak, $\Gamma_{\mathrm{low}}(T)$.

Figure 4 shows $I_{\mathrm{low}}(T)/I_{\mathrm{low}}(300\ \mathrm{K})$ along with the variation of the fraction of undenatured DNA determined by calorimetry. The two data sets are in remarkable agreement, proving that the intensities reflect the thermal denaturation of DNA.

The widths give the size of the closed domains of the DNA as a function of temperature, which is information that is inaccessible in conventional techniques for following the melting transition. The widths could be determined up to temperatures very close to the melting transition, however the Bragg peaks then become so small that the fit accuracy becomes unreliable. The large fluctuations observed in the immediate vicinity of the transition should not be regarded as significant. It is convenient to use reduced variables to compare the experimental widths with the theoretical analysis. The variables are $\Gamma_{\mathrm{low}}/Q^{\mathrm{c}}_{\mathrm{low}}$ and $\Gamma_{\mathrm{high}}/Q^{\mathrm{c}}_{\mathrm{high}}$, where $\mathrm{Q}^{c}_{\mathrm{high}}$ is

the center of the middle peak in the high-Q triplet. As shown by figure 4, the reduced widths are remarkably constant until the denaturation temperature. This indicates that, although the fraction of open base pairs grows smoothly as the transition temperature is approached, the size of the closed clusters stays sufficiently large for the width to be controlled by the inhomogeneities of the structure of the double helix rather than by the size of the regions that remain closed. This is an important observation, which puts a severe constraint on the models that can describe the denaturation transition of DNA.

3. Analysis

A Bragg peak is due to scattering from oriented DNA segments that have preserved their double helix structure. Therefore, in order to analyze the results we must first calculate the expected diffraction pattern of a finite segment of double-helical DNA. We then must use a statistical physics analysis to determine the distribution of sizes of the closed segments as a function of temperature and add their contributions to get the scattering intensity for each scattering vector of interest.

3.1. Structure factor for a segment of A-DNA

Far from the denaturing transition, the closed segments of DNA extend over hundreds of base pairs. Furthermore, a statistical averaging over the structural and thermal fluctuations is required to analyze the temperature dependence of the scattering pattern. Therefore an analysis based on an all-atom description is not feasible. Fortunately it is only necessary to calculate the intensity and width of our specific Bragg peaks, hence we may apply a substantially simplified calculation.

Any structure factor model for DNA must account for the difference in the scattering amplitude from each of the base pairs due to the difference in the orientation of the base pair with respect to its neighbours, and to variations in the sequence and to thermal fluctuations. These latter terms introduce distortions that affect the structure factor. Similar to our work on B-DNA [8], we adopted a simplified picture whereby the molecule was modeled by a chain of point scatterers at positions $\mathbf{r}_j$ and with scattering amplitudes of b_j. The structure factor from a closed segment, $S_m(\mathbf{Q})$, is expressed as

$$S_m(\mathbf{Q}) = \frac{1}{m}\left\langle \sum_{j_1=1}^{m}\sum_{j_2=1}^{m} b_{j_1}\, b_{j_2} \exp\Big[\, \mathrm{i}\mathbf{Q}\cdot\Big(\mathbf{r}_{j_1} - \mathbf{r}_{j_2}\Big)\Big]\right\rangle, \tag{5}$$

where $\langle\rangle$ designates a statistical average over both the thermal fluctuations and the variation in b_j and $\mathbf{r}_j$ due to the sequence. As for the calculation for B-DNA [8], the statistical average reduces to

$$S_m(\mathbf{Q}) = \sum_{j_1=1}^{m} b_{j_1}^2 + \frac{2}{m}\sum_{j_1=2}^{m}\sum_{j_2=1}^{j_1-1} b_{j_1}\, b_{j_2} \cos\Big[\, Q_{\|}\big(j_1 - j_2\big)a\Big] \times \exp\left\{-\frac{1}{2}\big(j_1 - j_2\big)\Big[\, Q_{\|}^2\Big(\left\langle\delta^2\right\rangle + \left\langle\sigma^2\right\rangle\Big) + Q_{\perp}^2\left\langle\eta^2\right\rangle\Big]\right\}, \tag{6}$$

where a is the average lattice spacing along the helical axis, $\langle\delta^2\rangle$ and $\langle\eta^2\rangle$ are the standard deviations of the actual positions of the base pairs with respect to a regular one-dimensional lattice in the longitudinal and transverse directions respectively. These variations correspond to the local distortions of the helical structure due to the sequence. The parameter η also accounts for the fact that, in A-DNA, the base pairs are shifted off axis. The standard deviation of the longitudinal thermal fluctuations, $\langle\sigma^2\rangle$, can be calculated from the stretching elasticity of DNA, deduced from the speed of sound in the molecule [16, 17]. We neglected the transverse thermal fluctuations because, as long as DNA stays in the helical form, they are much smaller than the off-axis structural fluctuations. Large transverse fluctuations imply that the corresponding segment is 'denatured' in the statistical analysis.

The unit cell that gives rise to Bragg peaks from A-DNA has 11 base pairs, corresponding to one full twist of its helicoidal structure. We can exploit the symmetry by applying a function to b_j with a periodicity of 11. At least one of the 11 values of b_j must differ from the rest, giving meaning to the unit cell, and $b_{\min} \approx 0.1 b_{\max}$ to approximate the effect on the intensity of the form factor from the shape of the molecule. Calculations for large closed segments can then be performed rapidly, testing the ability of our simple model to represent the real structure factor. Defining the structure factor for a single unit cell as $F(\mathbf{Q}) = S_{m=11}(\mathbf{Q})$, the structure factor $S_{N_{\text{cells}}}(\mathbf{Q})$ can then be calculated by summing over the different cells

$$S_{N_{\text{cells}}}(\mathbf{Q}) = \frac{1}{N_{\text{cells}}} \sum_{l_1=1}^{N_{\text{cells}}} \sum_{l_2=1}^{N_{\text{cells}}} F(\mathbf{Q}) \exp\left[\mathrm{i}\left(l_1 - l_2\right) a_{\text{cell}}\right] \times \exp\left\{-\frac{11}{2}\left|l_1 - l_2\right|\left[Q_{\|}^2\left(\left\langle\delta^2\right\rangle + \left\langle\sigma^2\right\rangle\right) + Q_{\perp}^2\left\langle\eta^2\right\rangle\right]\right\}, \tag{7}$$

where l_1, l_2 are the indices of the cells and $a_{\text{cell}} = 11\ a$ is the size of the unit cell. Note that the contributions of the fluctuations around the perfect lattice have a standard deviation multiplied by 11 because, for the positions of the cells, the variations are the sum of $m = 11$ individual variations. This is true both for the longitudinal displacements, for which the structural variations sum, and for the transverse contributions, where the lateral displacements lead to a curvature of the molecule which sums over cells. The double summation of equation (7) can be reduced to

$$S_{N_{\text{cells}}}(\mathbf{Q}) = N_{\text{cells}} F(\mathbf{Q}) + 2F(\mathbf{Q}) \sum_{s=1}^{N_{\text{cells}}} \left(N_{\text{cells}} - s\right) \cos\left(\mathbf{Q} s a_{\text{cell}}\right) \times \exp\left\{-\frac{11}{2} s\left[Q_{\|}^2\left(\left\langle\delta^2\right\rangle + \left\langle\sigma^2\right\rangle\right) + Q_{\perp}^2\left\langle\eta^2\right\rangle\right]\right\}. \tag{8}$$

Using equation (8) reduces the calculation time for equation (5) when m is large and is not a multiple of 11, so long as a correct summation over the remainder of base pairs is included.

We have tested various choices for the 11 phenomenological scattering amplitudes b_j in $F(\mathbf{Q})$. None of the choices precisely reproduce the expected structure factor of A-DNA, for which the dominant peaks are at positions 2 and 6,7,8 $\times 2\pi/(11\ a)$. This failure is not surprising because a simple 1D model cannot pretend to describe the complex 3D structure of A-DNA required to quantitatively calculate the relative intensities between different Bragg peaks. However, the widths of the peaks are only very weakly dependent on the choice of the b_j parameters, and for the calculation we eventually used the simplest sequence of 11 scatterers

with all but one of the b_j being equal. The model is useful because it enables us to calculate the widths of Bragg peaks from closed DNA segments with a very large variety of sizes, from a few units to thousands, and to include them in a statistical average. Despite its simplicity it reproduces the non-trivial experimental observation that the reduced widths of the Bragg peaks $\Gamma_{\mathrm{low}}/Q^{\mathrm{c}}_{\mathrm{low}}$and $\Gamma_{\mathrm{high}}/Q^{\mathrm{c}}_{\mathrm{high}}$ are almost equal if we set the standard deviations of the structural fluctuations to $\delta = 0.09$ Å and $\eta = 1.1$ Å. The value of δ is deduced from a structural analysis of a large variety of DNA sequences [18]. The same analysis gives an average of 0.495 Å for the transverse slide and shift of the base pairs, but does not account for the off-axis displacement of the base-pairs in A-DNA, which increases the contribution of the transverse fluctuations in the reduced 1D model that we use, hence our larger value of η.

3.2. Statistical analysis of the closed regions of DNA

The calculation of the size distribution of the closed segments of DNA is based on statistical physics analysis using the PBD model [19, 20] which is sufficiently simple to allow the analysis of DNA segments of tens of thousands of base pairs. The configuration energy of a DNA molecule of N base pairs is written

$$H_y = \sum_{j=1}^{N-1} W_j\left(y_j, y_{j+1}\right) + \sum_{j=1}^{N} V_j\left(y_j\right), \tag{9}$$

where y_j is the stretching of the jth base pair due to transverse displacements of the bases.

To describe the experimental observations on A-DNA, the model must be modified with respect to the version used for B-DNA [7, 8]. Figure 4 shows that the widths of the Bragg peaks stay almost constant even when DNA is partly denatured. This indicates that the size of the closed clusters stays so large that the widths are dominated by structural disorder rather than cluster size, even in the immediate vicinity of the transition. The investigation of some extensions of the PBD model [21] showed that large closed clusters are favored when the on-site potential $V_j(y_j)$ is not a simple Morse potential but includes a barrier for the reclosing of a base pair. The potential $V_j(y_j)$ is therefore written as [22]

$$V_j\left(y_j\right) = D_j\left(\mathrm{e}^{-\alpha_j y_j} - 1\right)^2 + \Theta\left(y_j\right)\frac{B_j y_j^3}{\cosh^2\left[c\left(\alpha_j y_j - d\,\ln 2\right)\right]}, \tag{10}$$

where $\Theta(y_j)$ is the Heaviside step function. The parameter B_j determines the amplitude of the barrier for closing, c its width, and d its position in units of the value of y_j at the inflexion point of the Morse potential.

The closing barrier has multiple origins. When a base pair is open, the exposed bases tend to form hydrogen bonds with water which must be broken for the base pair to close again. The structure of A-DNA, with its inclined bases and closely packed molecules in the film, also introduces some steric hindrance to closing. Moreover, in a mesoscopic model the 'potentials' represent the effect of many individual interactions and local fluctuations. They are actually free energies. As open bases have much more degrees of freedom than when they are stacked in the double helix, they have more entropy, which lowers the free energy of the open state. These different contributions are synthesized in the second term of $V_j(y_j)$.

Each parameter in the potential takes two different values depending on the type of base pairs. A–T pairs, in which the bases are linked by two hydrogen bonds, are easier to break than

Table 1. Potential parameters used in the statistical physics model (9). For the stacking interactions, the dimers are listed by giving the bases along one strand, in the 5′−3′ direction .

Potential V	D	α	B	c	d
A–T base pair	0.087 06 eV	3.0 Å^{-1}	6.00 eV	0.74 Å^{-1}	0.20
G–C base pair	0.088 30 eV	3.4 Å^{-1}	9.00 eV	0.74 Å^{-1}	0.20
Potential W		$\rho = 15$	$b = 0.8$ Å^{-1}		
Nucleotide stacking		AT, AA	AC, TC, AG, TG, CC, GG		
$k_{j,j+1}$(eV Å^{-2})		0.0041	0.0140		
$k_{j,j+1}(1+\rho)$(eV Å^{-2})		0.0656	0.224		

the G–C pairs, linked by three hydrogen bonds. We used parameters similar to those used to describe the melting curves of specific B-DNA sequences [22] introducing B, to account for the geometrical constraints in A-DNA, and lower values of D_j, giving shallower potentials consistent with the lower values of the melting enthalpies observed in our calorimetry studies.

The term $W_j(y_j, y_{j+1})$ describes the stacking interactions of the bases. We use the same functional form as in the standard PBD model

$$W\left(y_j, y_{j+1}\right) = \frac{1}{2}k_{j,j+1}\left[1 + \rho e^{-b\left(y_j + y_{j+1}\right)}\right]\left(y_j - y_{j+1}\right)^2, \tag{11}$$

which accounts for the weakening of the interactions when the pairs are broken [19, 20]. A uniform stacking could be assumed for B-DNA [8]. For A-DNA it was necessary to account for the dependence of the stacking interaction, $k_{j,j+1}$, on the sequence to describe properly the broader melting transition and the large closed clusters.

There are ten independent permutations for how base pairs may stack in DNA, and hence ten possible types of stacking interaction [22]. However, we could describe melting curves to a good accuracy with *only two different interaction constants*: one when A–A or A–T are stacked; and one for all other possible combinations. The stacking interactions were chosen to properly describe the melting profile of our DNA sample, as determined by calorimetry.

Figure 1 shows a comparison of the experimental melting profile with the theoretical profile calculated with the modified PBD model. Some fine structure is observed in the calculated profile due to the finite number of base pairs used for the calculation. The experimental data were obtained from a sample with billions of base pairs, while the PBD model used a segment of 40 000 base pairs from the sequence of one salmon gene. The melting profile is, of course, sequence dependent, and we expect that the fine structure in the calculation would average out if we could perform the calculation for a much larger DNA segment. Nevertheless, the model gives a denaturation profile which is very close to the calorimetric profile.

The parameters chosen for our model are summarized in table 1. It is useful to compare these parameters with those of B-DNA [7, 8] although, because we have two stacking interactions for A-DNA as opposed to one for B-DNA, a fully quantitative comparison is not

possible. The value for ρ affects the sharpness of the transition. We chose $\rho = 15$ for A-DNA which, combined with the inhomogeneous stacking interactions, gives a much broader transition than for B-DNA with $\rho = 50$. The effective coupling constant between closed base pairs is given by $k_{j,j+1}(1 + \rho)$. We expect this to be stronger in A-DNA as the distance between base pairs is smaller than in B-DNA, hence we were obliged to increase $k_{j,j+1}$ to compensate for the decrease in ρ. For B-DNA, the product $k_{j,j+1}(1 + \rho) = 0.022\,95$ eV Å^{-1} is smaller than both the values for A-DNA listed in table 1.

The model also allows us to calculate the probability, $P(m, T)$, of having a closed cluster of size m at temperature T [8]. The structure factor of the DNA sample is obtained by averaging over all the possible cluster sizes

$$S(\mathbf{Q}, T) = \sum_{m=1}^{\infty} P(m, T) S_m(\mathbf{Q}, T). \tag{12}$$

The summation is computed exactly up to cluster sizes $M = 160$ and then asymptotic forms of $S_m(\mathbf{Q}, T)$ and $P(m, T)$ are used to extend it for larger clusters [8]. The calculated structure factor $S(\mathbf{Q}, T)$ is then fitted by the same expression as the experimental data (equation (1)) for a direct comparison between theoretical and experimental intensities and widths.

4. Discussion

Figure 4 summarizes the result of the analysis and its comparison with the experimental data. The temperature variations of the fraction of denatured base pairs obtained from calorimetry, from the statistical physics analysis, and from the normalized intensity of the neutron Bragg peaks are plotted. As previously noted, there is satisfactory agreement between the two experimental quantities. The calculated fractions also agree, which is unsurprising as the parameters were carefully chosen to reproduce the melting curve as given by calorimetry.

More importantly, figure 4 also shows a comparison between the variation of the observed and calculated reduced widths of the Bragg peaks. The experimental observation that $\Gamma_{\mathrm{low}}/Q^{\mathrm{c}}_{\mathrm{low}} \approx \Gamma_{\mathrm{high}}/Q^{\mathrm{c}}_{\mathrm{high}}$ is correctly reproduced in the theoretical analysis. There is a small discrepancy between theory and experiment, with the theory predicting a slight increase in the width with temperature for $\Gamma_{\mathrm{low}}/Q^{\mathrm{c}}_{\mathrm{low}}$. This indicates the limits of a simplified model that uses only the base pair opening to describe the complex temperature-dependent structure of A-DNA. Despite the discrepancy, there is satisfactory quantitative agreement.

The theory predicts, and the experimental data show, that the closed regions of A-DNA maintain a long spatial coherence, not dropping below ~90 base pairs until the very large stage of the transition. The reduced widths are thus almost independent of temperature until very close to the transition. Our previous study on B-form DNA arrived at similar conclusions, suggesting that this might be a generic feature of DNA melting irrespective of the conformational structure.

The insert of figure 1 shows the calculated average size of the closed domains of A-DNA. Surprisingly, the calculation shows a *sharp rise* in the average length of a closed region just before complete denaturation, also seen as a dip in the calculated widths in figure 4. This indicates that those few closed regions that remain at these temperatures must be quite long indeed. The precise calculated lengths are sequence-specific, however a sharp rise was observed for all the salmon sequences we tested and comes from extended GC-rich regions of the genome

which are the last to denature. Although the poor signal-to-noise in the neutron data in the very last stages of denaturing makes reliable fitting difficult, there is a systematic trend that the reduced widths fall in the same temperature region as the model predicts. We believe that this is a real effect and is a further vindication for the model.

No such sharp rise was predicted for B-form DNA. However, the presence of such a rise cannot be ruled out as the collapse of the fiber structure prevented the neutron experiment from following the DNA structure through the full transition.

The difference between the expressions for the potentials $V_j(y_j)$ required to successfully model the data for A- and B-DNA is worthy of note. A simple Morse potential sufficed to model the B-DNA data, while for the A-form it was necessary to include a barrier between the 'open' and 'closed' positions for a nucleotide. Previous studies of the dynamics of DNA in B-DNA have shown that a similar barrier was required when calculating the time a base pair stays open after denaturing [21], however this appears to be unnecessary when calculating the static correlation lengths measured using neutron scattering. The need here to include the barrier for the A-form calculations hints at important differences between the dynamics of the two confirmations, having ramifications on the stability of the molecule and the ability to transcribe information from this conformation.

5. Conclusion

The fiber structure of A-DNA in a spun film does not 'collapse' when the double helix denatures under the influence of thermal fluctuations. We were able to follow the complete helix-denaturing transition from pure A-DNA films using calorimetry and neutron scattering. The DNA thermal denaturation is much smoother than for the B-form, suggesting a weakly first order transition with stronger coupling between base pairs and thus a stiffer chain. Moreover the data show that the domains that diffract coherently stay large (several turns of the double helix) even when the molecule is almost fully denatured.

The PBD model, previously tested for melting profiles and neutron scattering of B-DNA samples, was adapted to quantitatively describe A-DNA by modifying the on-site potential to include a barrier for the reclosing of the base pairs, and by taking into account the sequence dependence of the stacking interactions. The model can be kept fairly simple by introducing only two values of the stacking interactions instead of the possible 10.

Acknowledgements

Part of the numerical calculations have been performed with the facilities of the Pôle Scientifique de Modélisation Numérique (PSMN) of ENS Lyon. We thank ANSTO and the ILL for the use of the neutron instruments, along with their technical staff.

References

[1] Watson J D and Crick F H C 1953 *Nature* **171** 737
[2] Fuller W 2003 *Nature* **424** 876
[3] Franklin R E and Gosling R G 1953 *Acta Crystallogr.* **6** 673
[4] Cheetham G M T and Steitz T A 1999 *Science* **286** 2305

[5] Mikhailenko I A and Shlyakhtenko L S 1984 *J. Biomol. Struct. Dyn.* **1** 1501
[6] Sebastiani F, Pietrini A, Longo M, Comez L, Petrillo C, Sacchetti F and Paciaroni A 2014 *J. Phys. Chem.* B **118** 3785
[7] Wildes A, Theodorakopoulos N, Valle-Orero J, Cuesta-López S, Garden J-L and Peyrard M 2011 *Phys. Rev. Lett.* **106** 048101
[8] Wildes A, Theodorakopoulos N, Valle-Orero J, Cuesta-López S, Garden J-L and Peyrard M 2011 *Phys. Rev.* E **83** 061923
[9] Rupprecht A 1970 *Biochim. Biophys. Acta* **199** 277
[10] Blake R D and Delcourt S G 1998 *Nucleic Acid. Res.* **26** 3323
[11] Tan Z J and Chen S J 2006 *Biophys. J.* **90** 1175
[12] Stellwagen E, Muse J M and Stellwagen N C 2011 *Biochemistry* **50** 3084
[13] Valle Orero J, Wildes A, Garden J-L and Peyrard M 2013 *J. Phys. Chem.* B **117** 1849
[14] Wilkins M H F, Stokes A R and Wilson H R 1953 *Nature* **171** 738
[15] Valle-Orero J, Garden J-L, Wildes A, Richard J and Peyrard M 2012 *J. Phys. Chem.* B **116** 4394
[16] Krisch M, Mermet A, Grimm H, Forsyth V T and Rupprecht A 2006 *Phys. Rev.* E **73** 061909
[17] van Eijck L, Merzel F, Rols S, Ollivier J, Forsyth V T and Johnson M R 2011 *Phys. Rev. Lett.* **107** 088102
[18] Lavery R, Moakher M, Maddocks J H, Petkeviciute D and Zakrzewska K 2009 *Nucleic Acid Res.* **37** 5917
[19] Peyrard M and Bishop A R 1989 *Phys. Rev. Lett.* **62** 2755
[20] Dauxois T, Peyrard M and Bishop A R 1993 *Phys. Rev.* E **47** R44
[21] Peyrard M, Cuesta López S and James G 2009 Nonlinear analysis of the dynamics of DNA breathing *J. Biol. Phys.* **35** 73–89
[22] Peyrard M, Cuesta López S and Angelov D 2009 Experimental and theoretical studies of sequence effects on the fluctuation and melting of short DNA molecules *J. Phys.: Condens. Matter* **21** 034103
[23] Quinn N L *et al* 2008 *BMC Genomics* **9** 404 GeneBank EU481821