Confinement, spatial correlations and flexibility in the melting transition of DNA

University of Burgos and Institut Laue-Langevin

Adrián González Rodríguez

Advisors: Andrew Wildes and Santiago Cuesta López

December 20, 2017

Contents

Chapter 1

Introduction

Deoxyribonucleic acid (DNA) is a macromolecule which carries the genetic information essential for every living organism. It is a biopolymer built by the repetition of monomers called nucleotides. Each nucleotide consists of a phosphate group, a sugar group and a nitrogen base. The phosphate and sugar groups alternate in sequence and form a linear chain. Two of these chains are linked by hydrogen bonds between nitrogen bases generating the famous double helix. Each couple of linked nitrogenous bases is called a base pair.

The melting transition (thermally induced helix-to-coil transition) of DNA is a first order phase transition induced in DNA molecules by heat. In such a process the hydrogen bonds between base pairs break as the temperature is increased. Local openings of consecutive base pairs can occur at a temperature below the transition resulting in zones of open base pairs (called denaturation bubbles) which are surrounded by closed zones in which the double helix is intact. As the temperature is raised and more base pairs break, the open zones expand to larger sizes and eventually the two strands of the double helix separate. Fig. 1.1 illustrates the process.

The melting transition of this life-essential molecule has been heavily studied to learn about the intramolecular interactions, the impact of the base pair sequence in DNA unwinding and the effect of the solvent in DNA stability [2, 3, 4]. A better understanding of the

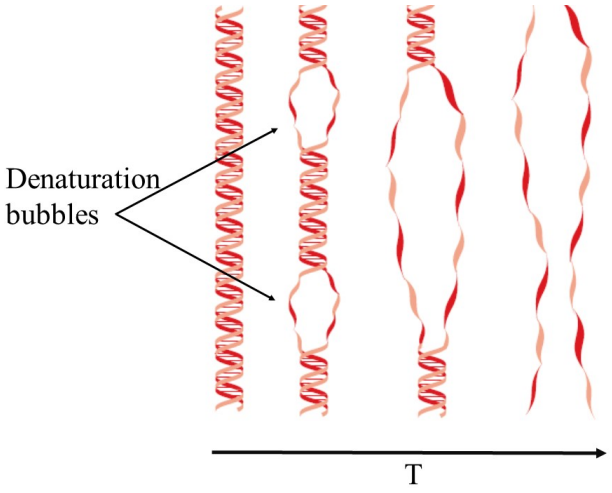


Figure 1.1: DNA openings as T increases. Adapted from [1].

transition could have an immediate impact in applications such as high resolution melting analysis or polymerase chain reactions, which are widely used in bio-labs around the world [5, 6]. However, the melting of DNA itself is interesting to study from a theoretical point of view since it is, essentially, a phase transition in a one dimensional system.

In this thesis, a better understanding of the melting transition of DNA and its effects in the functionality of the molecule is attempted by following two avenues: i) the study of highly oriented DNA fibers with wide angle diffraction techniques and calorimetry. ii) the investigation of a short chain, biologically relevant, sequence with small angle scattering techniques (SAS).

Highly oriented DNA fibers

Experimentally, the transition has been studied with techniques that either probe the bulk sample, like calorimetry, UV-Vis absorption and circular dichroism spectroscopy, or localized points on the molecule, like fluorescence spectroscopy. The spatial structure of DNA molecules is not accessible using these techniques, and therefore, the distribution of open and closed base pairs during the phase transition can not be determined. Knowledge of

the spatial correlations and how they evolve with temperature is necessary for a complete understanding of the transition.

Scattering techniques allow to access structural information and thus the spatial correlations within the molecule. Such information is more easily accessible when the sample has long-range order, which gives rise to Bragg peaks. DNA fibers, in which the molecules are coaligned and are close packed, have sufficient long-range order to exhibit Bragg peaks in their diffraction patterns. By studying these peaks, the information of interest is collected in a measurement which is only weakly perturbed by sample imperfections and incoherent contributions to the scattering.

The melting transition of oriented DNA fibers equilibrated with a humid atmosphere has been previously studied [7, 8] using neutron scattering. The results have been successfully modeled with the mesoscopic statistical-mechanical Peyrard-Bishop-Dauxois (PBD) model [9]. Therefore, these works were able to access the spatial correlations of the molecules during the transition and reinforce the validity of the PBD model, which is already widely used to describe complex DNA denaturation curves. During the analysis, the question of whether the confinement of the DNA molecules in the fibers has an effect on the transition arose. The restricted degrees of spatial freedom due to the confinement of the molecules could arguably have an impact on the interpretation of the results since the model does not account for intermolecular interactions.

In order to test the effects of the molecular confinement, a new experiment was designed in which the fibers were submerged in a solution with a given osmotic pressure. The osmotic pressure method has been reported in numerous articles [10, 11]. It consists of submerging the DNA fibers in a saline solution with a high-molecular-weight polymer such as polyethylene glycol (PEG). If the PEG does not penetrate the DNA fibers, the osmotic pressure acts analogously to a mechanical and permeable piston preventing DNA from dissolving, thus preserving the long-range order, while allowing water and salt to be exchanged between the fibers and the solution. X-ray diffraction has proved that fibers equilibrated with these kind of solutions show a well-defined interaxial distance between molecules, which increases with

decreasing concentration of added polymer [11].

A pilot experiment on fibers submerged in PEG solutions has been performed and reported previously [12]. It proved that it is possible to follow the DNA melting transition using scattering techniques on a submerged sample but the quality of the data was insufficient for a reliable comparison to a theoretical model.

Another avenue to modify the interaxial distance of the molecules in the fibers is to submerge them in ethanol-water mixtures. DNA is not soluble in these mixtures (for ethanol concentrations down to at least 20% v/v) therefore the fibers might swell and the long range orientation is preserved. Using two kinds of solutions which are so different in nature may help detect specific effects of the PEG in the transition which are not related with confinement.

In addition, the ethanol will modify the hydrophobic forces which are closely related to the base pair stacking. Therefore it could be instrumental to investigate the relation between base stacking and structural correlations in the melting transition.

The aim of the study of the melting transition of highly oriented DNA-fibers submerged in PEG and ethanol solutions was to understand better the dynamics of the transition by adding structural information which is mostly lacking and study the effect of the confinement created by the close packing of the DNA molecules on the melting.

Short chain DNA, the Widom sequence

Flexibility and bending of DNA play a important role in many biological processes. They are central for the packing of the genome inside the cell's nucleus. For example, in mammalian cells, approximately 2 m of linear DNA must be fit in a nucleus of diameter 10 μm. An important step in this extreme packing is performed by winding the DNA around eight histone protein cores forming what is called a nucleosome. Nucleosomes are later folded in several higher order structures to finally build the chromosome. Some sequences have shown a strong affinity to bind to the histones and their artificial synthesis has created the

possibility of studying the structure and parameters of the binding in the nucleosome core particle (NCP) [13]. One of these strong positioning sequences was chosen for the small angle scattering experiments described in this document. It is the so called Widom-601.

Despite bending of DNA being patently important the origin and even magnitude of its flexibility is still a controversial topic [14]. The melting transition could be closely related with the changes in the flexibility of DNA as a function of temperature. Thermally-induced opening of base pairs seems to be an essential factor to predict the temperature dependence of DNA flexibility [15].

The flexibility and geometrical conformations of a short DNA chain containing the Widom-601 sequence in solution at room temperature and during its melting transition were studied using small angle X-ray scattering and small angle neutron scattering (SAXS and SANS respectively). The aim for the room temperature study was to search for possible intrinsic mechanical and/or geometrical properties of DNA which may have a significant role in the wrapping of DNA around the histones. The temperature-dependent study aimed to acquire experimental data able to defend the hypothesis that local melting of base pairs is the main driving effect changing the flexibility of DNA as temperature raises.

Chapter 2

Methods and theory

The aim of this chapter is to familiarize the reader with the main experimental techniques used in this thesis and the theory behind them. The experimental protocols, data treatment, and sample preparation will be also reported here.

2.1 Sample preparation

The intrinsic characteristics of the samples used in this work were essential for accessing the information of interest. So a precise preparation of the samples was of capital importance for the success of the project. Next sections present the different methods and recipes used.

2.1.1 Production of oriented DNA fibers

For the present work, highly oriented DNA fibers were created using the wet spinning technique developed by Allan Rupprecht. The wet spinning method consists of precipitating DNA in a salty solution with high ethanol concentration, then gathering different threads into a single fiber which is stretched and then wound around a cylinder to form a film of coaxial fibers. Within each fiber, there are regions in which the DNA molecules are oriented

along the fiber axis. There exists also a periodic positioning of the molecules in the direction perpendicular to the fiber axis.

The technique and equipment was described in detail in the literature ([16, 17, 18]).

The starting material for the spinning is DNA with sodium salt from salmon testes purchased from Sigma-Aldrich company. The G-C content is approximately 41.2%. The molecular mass of this DNA was measured with gel electrophoresis. It was of the order of 20 kilo-base pairs with great polydispersity [7]. The initial solution used in the wet spinning method was prepared by dissolving the DNA in pure water with sodium or lithium counter ions. A 0.15 M $NaCl$ solution was used for making Na-DNA fibers, consisting of Milli-Q water and 1.58 g/L of DNA, 0.15 M NaCl, 0.003 M Na-citrate and 5×10^{-5} M Na-EDTA. For making Li-DNA fibers a 0.3 M $LiCl$ solution of 1.58 g/L DNA, 0.3 M $LiCl$, 0.2 g/L of sodium azide and 0.02 g/L of EDTA was used. The concentration of the counter-ion ($LiCl$) was doubled with respect to the $NaCl$ case to ensure a better exchange with the Na present in the raw product.

The wet Spinning apparatus

Fig. 2.1 presents an schematic of the wet spinning device. The glass vessel, glass column and the degasser were filled with a salt-water-ethanol (EtOH) mixture called the spinning bath. For Na-DNA the spinning bath was a 75% EtOH solution and 0.02 M $NaCl$. In the case of Li-DNA it was a 83% EtOH solution at 0.4 M $LiCl$. The DNA solution was degassed under vacuum for 4 hours before starting the spinning since the formation of bubbles during the spinning process can ruin the sample. The spinneret is a ceramic overture with many small holes (like a shower head, fig. 2.2a); the degassed DNA solution was introduced in a syringe and extruded through the spinneret into the top of the glass column at a rate of 33 ml/h. This guaranteed a laminar flow of the DNA solution into the column. The DNA precipitates at a certain distance below the spinneret. It then falls down the column as several fibers before being collected into a single fiber at a V-shaped teflon (polytetrafluoroethylene) coated wire at the botton of the column and in the glass vessel. This V-shaped wire guides the

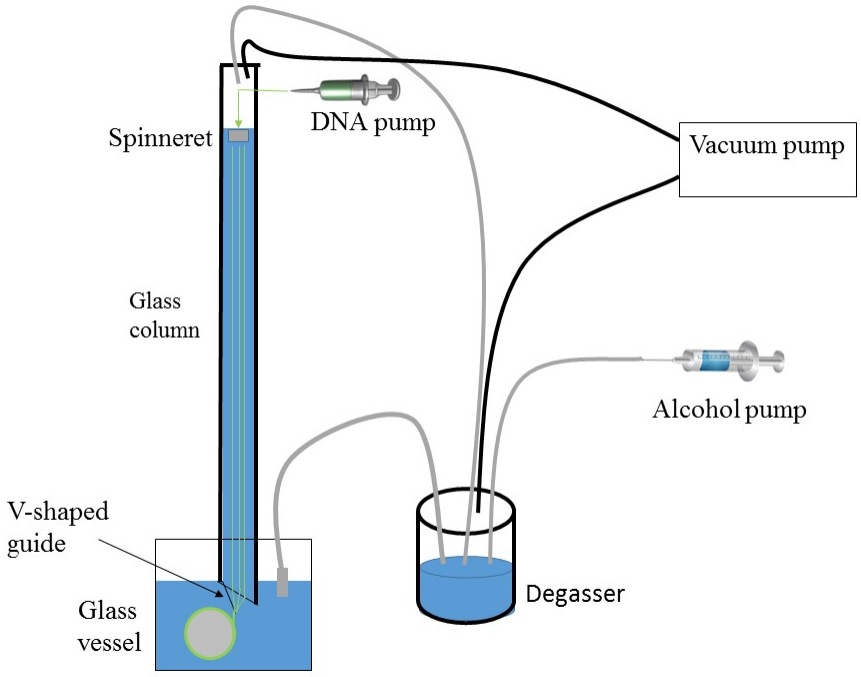
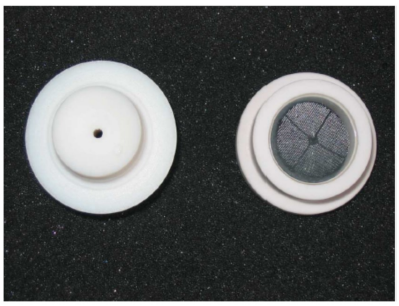


Figure 2.1: Scheme of the wet spinning apparatus.

resultant singular fiber onto a rotating teflon-coated cylinder (fig. 2.2b). The fiber is wound around the cylinder which also undergoes a linear translation. The width of the resulting film can be controlled by adjusting this translation. Pure ethanol was injected constantly in the degasser and the degasser solution was pumped to the top of the column in order to maintain the concentration of EtOH in the column despite the injection of the water-base DNA solution. Since the initial volume of ethanol is approximately 3 to 1 of water, a good rate of injection of ethanol in the degasser is around 100 ml/hour, i.e., three times faster than the water-based DNA solution is injected. A recycling cycle was carried out each 10 minutes. During this cycle the vacuum pump was activated and spinning bath was moved from the glass vessel to the degasser. These cycles prevented the glass vessel from overflowing and the degasser from emptying.

The wet spinning apparatus is an almost unique device that was donated to the ILL in 2004 by H. Grimm. It was installed in the bio preparation lab by Mark Johnson and Fabian Fontaine-Vive. Ammandine Vittoz prepared an interface for the computer control of the



(a)

(b)

Figure 2.2: a) Two teflon parts that form the spinneret. b) Teflon coated cylinder.

different parts through Labview.

Film treatment

Once spinning is completed, the final DNA film is obtained after the cleaning, drying and re-humidification processes of the viscous film on the cylinder. During the cleaning process the cylinder with the spun fibers was submerged in a salt-water-ethanol solution under gentle stirring for 24 hours in order to wash away salt excesses and impurities. The cleaning solution used for $Na-$DNA fibers was a 77% EtOH solution and 0.03 M $NaCl$. For Li-DNA films it was a 80% EtOH solution and 0.25 M $LiCl$. The cylinder was afterwards inserted in a desiccator with silica gel for slow drying during 24 hours. In this drying step the excess water and alcohol were removed and the actual film was formed. The final step before the removal of the film from the cylinder was a re-humidification procedure in which the cylinder was introduced in another desiccator with an over-saturated salt solution to control the relative humidity.A supersaturated $NaCl$ solution was used for equilibrating the DNA film, giving 75% relative humidity. This process stabilized the structure of the film and made the film removal easier. A picture of the resultant films can be seen in fig. 2.3.

Figure 2.3: Picture of the final films produced by the wet spinning apparatus.

Submerged fibers

$Na-$ and Li-DNA oriented fibers submerged in PEG solutions made with ^2H_2O (heavy water) and ethanol/^2H_2O mixtures have been investigated in this work using calorimetry, X-ray and neutron scattering. Using heavy water is an advantage when investigating samples with neutrons because it minimizes the incoherent scattering [12]. ^2H_2O was used in the samples studied with other techniques for consistency. Since the chemical properties of heavy water and normal water are almost identical the range of application of the results is not hindered by the use of heavy water.

The PEG solutions were made by dissolving the specific amount of PEG in a buffer made with ^2H_2O with 10 mM Tris (tris(hydroxymethyl)aminomethane), 1 mM EDTA (ethylene-diaminetetraacetic acid). Unless mentioned otherwise the solutions contained 0.1 M $NaCl$ for samples made with Na-DNA and 0.1 M $LiCl$ for samples made with Li-DNA. The PEG solutions will be identified by the amount of PEG in weight percent.

The ethanol/water mixtures were prepared with the same buffer and will be identified by the amount of ethanol in volume percent. Also, for the sake of simplicity the mixtures will be addressed as ethanol solutions. For the neutron samples fully deuterated ethanol was used (ethanol-d6), for the other techniques protonated ethanol was used due to the cost of

ethanol-d6.

A specific sample for neutron scattering consisted of a number of films with a total DNA mass close to 0.6 g that were concertina folded and stacked with their axes coaligned. The dimensions of the final sample were approximately $2.5 \times 2.5 \times 0.2\ cm^3$. The DNA was placed in a niobium envelope which in turn was placed inside an aluminum cassette. The niobium is necessary since the DNA can react chemically with aluminum at high temperatures. Around 0.8 ml of a given PEG or ethanol solution was degassed and placed in the cassette so the DNA was totally submerged. Then the cassette was screwed closed and sealed with a lead-wire gasket.

2.1.2 Widom-sequence

The short DNA sequence chosen for the small angle scattering experiments described in this document is the so called Widom-601 sequence. It has great affinity to bind around protein histone octamers and it is called a strong positioning sequence. It was discovered by Lowary and Widom [19]. It has 145 base pairs with a total length of around 485 Å. Fig. 2.4 presents the primary structure of the sequence with the binding site highlighted in red.

The artificial DNA needed for the experiments was synthesized by collaborators in the ICCRAM at the university of Burgos, Spain (Marta Marty Roda, Lorena Romero Santacreu), and in the INMG at the university of Lyon, France (Ramachandran Boopathi, Dimitar Angelov). The exact protocol differed slightly from one lab to another but a general summary

CAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCC
CGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAG
GCACGTGTCAGATATATACATCGAT

Figure 2.4: The 145 base-pair sequence corresponding to the NCP-601 nucleosome investigated in [13] also known as Widom-601 sequence. The fragment highlighted in red is the strong positioning element characteristic of this sequence.

is presented below.

The starting material was a solution of plasmid pGEM-3z/601 purchased from Addgene, which contains the Widom sequence. A polymerase chain reaction (PCR) was performed with the proper primers in order to duplicate the Widom sequence. After enough duplications, the result was processed with a purification column kit (Gene Matrix Basic) for separating the Widom sequences from the remanent plasmid. PCR alone is not an optimal method to obtain the quantities of DNA needed for neutron scattering (in the order of the mg) but the amount of widom-DNA out of the first PCR allowed to use another multiplication technique, called competent cells. In cell biology, competence is defined as the ability of a cell to take extracellular DNA and incorporate it in its vital cycle, including reproduction. Thus, for reproducing a given sequence which has been assimilated by a cell culture it is sufficient to allow the cell culture to grow. Cells will not assimilate a short sequence such as the Widom-601 in this way but they can assimilate and multiply a plasmid including the Widom sequence. Therefore the Widom DNA from the first PCR and purification was used in a ligation process with ligase T4 in which plasmids pGEM/3z-601 were reformed. The plasmids were introduced in competent cells DH5α. After adequate growing time the plasmids were extracted from the cells by GeneJET Plasmid Miniprep Kit from Thermo Scientific. Finally, repeating the PCR and column purification stated above and performing further purification using ethanol/chloroform/fenol extraction as well as precipitation via centrifugation allowed the desired quantities of Widom-sequence DNA molecules to be obtained.

2.2 Differential scanning calorimetry

Differential scanning calorimetry (DSC) is a thermo-analytical technique designed to gauge the change in specific heat of a sample as a function of temperature. Many thermodynamical quantities may be calculated from calorimetry data including transition enthalpies, melting temperatures, specific heat and free energy. Apart from the references specifically highlighted

in the text, general information for the elaboration of this section has been collected from the book by Hohne [20].

2.2.1 Thermodynamical principles

A given conformational or phase transition can be described by the change in the thermodynamical properties: internal energy (ΔU), enthalpy (ΔH), Gibbs free energy (ΔG), Helmholtz free energy (ΔF) and entropy (ΔS). When the pressure is constant the enthalpy is a quantity equivalent to the total heat content of a system. The entropy represents the amount of energy in a system which is unavailable for performing work. It also is understood as the degree or disorder within the system [21, 22].

If the pressure is kept constant the difference in enthalpy and entropy during a transition induced by a change in temperature can be calculated from the specific heat (C_p):

$$(\Delta H)_p = \int_{T_1}^{T_2} C_p dT \tag{2.1}$$

$$(\Delta S)_p = \int_{T_1}^{T_2} \frac{C_p}{T} dT \tag{2.2}$$

Where the specific heat is defined as the amount of heat per unit mass needed to raise the temperature of the system by one degree at constant pressure.

2.2.2 DSC technique

There are two types of DSC experiments: the power-compensated DSC; and the heat flux DSC which is the technique of choice for the present work. In a heat flux differential scanning calorimeter two thermally isolated cells, one for the sample under study and another for the reference, are placed in a furnace connected to a heater. If the temperature of the furnace is raised the temperature of the sample and the reference cell will differ because they contain

samples with different C_p. The difference in temperature relates to the differential heat flux through K_e, the heat exchange coefficient of the furnace: $\Delta P = K_e \Delta T$. K_e can be obtained with a calibration using a material of well known thermal properties.

The differential heat capacity of the sample with respect to the reference can be expressed in function of the differential heat flux (derivation can be found in [23]):

$$\Delta C_p = -\frac{\Delta P}{\beta} - \tau_{ins} \frac{d(\Delta P)}{dT} \tag{2.3}$$

where $\beta = \frac{dT_0}{dt}$ is the scanning rate of the experiment (T_0 is the furnace temperature) and τ is the so called thermal time constant of the calorimeter [24]. This constant is related with the thermal lag of the set-up of the instrument and can be approximated by controlling the time it takes for the instrument to respond to a swift power change.

Figure 2.5 is an example of a calorimetry scan presenting heat flux (J/s) as a function of temperature during a heating ramp. If the sample does not undergo a physico-chemical transition in a given range of temperatures then the heat flux into the sample is constant in this range. On the contrary, when a thermal process takes place within the sample the heat flux changes and the sign of this change depends on whether the process is exothermic, like the crystallization transition of the figure, or endothermic, like the melting. In the former case heat is produced by the sample during the transition and so a lower heat flux is going into the sample. The opposite is true for the latter case.

2.2.3 Instruments

The differential scanning calorimeter used for collecting the data of the fiber samples of the present work was a MicroDSC III from Setaram instrumentation. It has two 1 ml volume Hastelloy C (Nickel and Chromium alloy) cells, the sample cell and the reference. The cells are placed in highly conductive tubes, thermally isolated from each other and from the outside environment. Thermocouples are fixed to the tubes for temperature measurement.

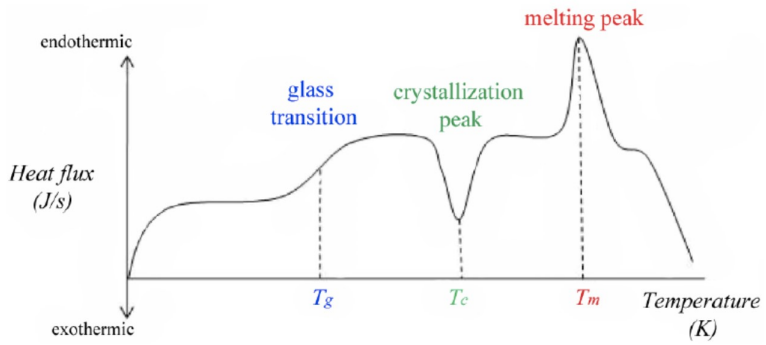


Figure 2.5: Example of differential scanning calorimetry scan. Three different transitions can be observed: a glass transition, a crystallization and a melting. reproduced from [25].

Figure 2.6: MicroDSC III from Setaram instrumentation company. Reproduced from [27].

During a scan both cells are subjected to a temperature ramp. Constant circulating toluene allows for a very precise temperature stability. The cooling is achieved through an external stabilization water bath which was kept at 27 °C for all experiments. The temperatures ranged from room temperature to 120 °C and the programmable scanning rate is 0.001 to 1.2 °C/min. The thermal time constant (τ_{ins}) as provided by the manufacturer is 60 s [26, 27].

A sample of dissolved DNA was measured in a different calorimeter from the rest of samples (a Nano DSC III from Calorimetry Sciences corp.) due to sensitivity reasons. The DNA concentration of the dissolved sample was 2 mg/ml and the volume of solution measured was 0.33 ml. The working principle of this calorimeter is the same as for the MicroDSC III and the methods used are identical. The sample was measured by Hervé Guillou in the

Institut NÉEL, Grenoble, France.

2.2.4 Experimental protocol

Most of the calorimetry work performed for the thesis was on samples of highly oriented fiber Na/Li-DNA submerged in PEG solutions or ethanol/^2H_2O mixtures described in 2.1.1. For the sake of convenience the liquid in which a given fiber sample was submerged would be referred to simply as solution in this section.

In a typical experiment the DNA amount under study was a piece of film that weighted around 25 mg when dry. The solution used for the experiment was degassed using ultrasound for 20 min and then 0.5 ml was used to submerge the DNA. The degassing minimized the quantity of bubbles introduced during the process since they may increase the noise of the signal or generate unstable baselines [28]. The DNA was left to equilibrate with the aqueous medium for around a week before being measured.

The scanning rate β was fixed at 1 °C/min for all the experiments. Increasing it to 1.2 °C/min or decreasing it to 0.5 °C/min made no significant difference in the data recorded.

Before the DNA sample was measured, a "solution versus solution" scan was performed where 0.5 ml of degassed buffer was place in both sample and reference cell. The thermal program for these scans consisted of successive cycles of heating and cooling from room temperature to around 110 °C. The calorimetric curve evolves during the first cycles but eventually stabilized to a constant value of around zero. The instability of the curve in the first scans is a well known instrumental artifact related to the fact that there is a natural instrumental baseline for a particular set of scan conditions which disappears after several heating cycles [29]. The curve of the heating ramp of the last cycle was reproducible and it was taken as the reference curve for the DNA sample and used in the data analysis. Thus, exactly the same conditions of the solution-solution scan must be used for the sample in order to avoid the reappearance of this artifact.

After the reference curve was obtained the sample (DNA+solution) was introduced in the sample cell and the DNA melting experiment was launched, with the same 0.5 ml of buffer in the reference cell. Since approximately the same amount of buffer (0.5 ml) is in both cells, features in the final curve could be ascribed only to the internal DNA or the DNA-buffer interactions.

2.2.5 Data analysis

The raw output data of a heat flux DSC device is a curve of differential heat flux in power units versus temperature ($\Delta P(T)$), c.f. fig. 2.5. The aim of this subsection is to explain how to use these data to calculate thermodynamical quantities that are relevant to characterize a transition. Quantities of choice are the melting (or denaturation) temperature, the width of the transition, and the fraction of open base pairs. The final methodology used for data analysis as well as most of the content of this section have been heavily based on several works of Charles H. Spink [29, 30].

The reference curve was subtracted from the measured $\Delta P(T)$. Then $\Delta P(T)$ can be converted to $\Delta C(T)$ in J/°C via eq. 2.3. The differential heat capacity per unit mass ($J/g°C$) can be obtained by:

$$\Delta c(T) = \frac{\Delta C(T)}{m_{DNA}} \tag{2.4}$$

where m_{DNA} is the mass of the DNA in the sample cell. The resulting curve usually has a single peak over a straight-line baseline. T_{min} and T_{max} are respectively the minimun and maximum temperatures of the scan. Before further treatment a baseline was defined as the connection line between $\Delta c(T_{min})$ and $\Delta c(T_{max})$ and then subtracted from the data. This curve could be fitted with a Gaussian function and so the full width at half maximum (FWHM) could be determined. This parameter was taken as the width of the transition.

The differential heat capacity allows the calculation of the fraction of base pairs which

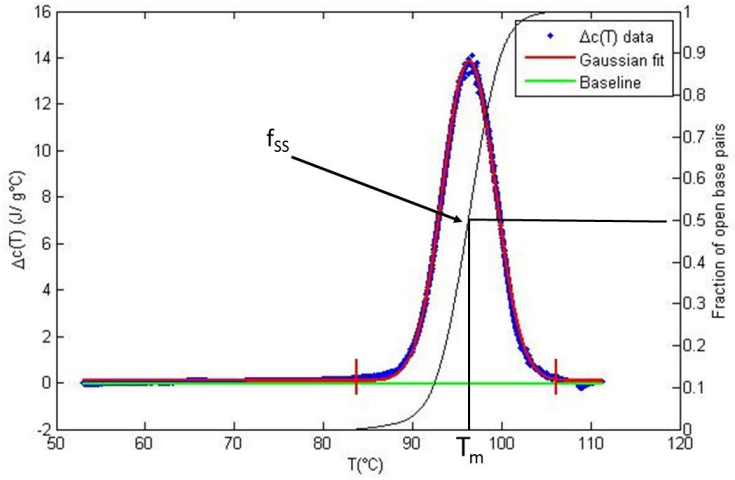


Figure 2.7: Example of data reduction and analysis performed on the calorimetric curve of a sample of Na-DNA submerged in a PEG solution. Left y-axis: Δc in blue, Gaussian fit to the data in red and baseline in green. Right y-axis: Fraction of open base pairs in black. The melting temperature is highlighted.

have opened at a given temperature: f_{SS}, where SS stands for single strand. $f_{SS}(T)$ was calculated as done in [31] with the equation:

$$f_{SS}(T) = \frac{(\Delta H)_n(T)}{\Delta H} \tag{2.5}$$

Where $(\Delta H)_n = \int_{T_{min}}^{T} \Delta C_p(T) dT$ is called the cumulative enthalpy at temperature T. Once the fraction of open base pairs is known the fraction of base pairs which are closed at a specific temperature is calculated as:

$$f_{DS}(T) = 1 - f_{SS}(T) \tag{2.6}$$

where DS stands for double strand. The melting temperature, T_m is the temperature at which half of the base pairs are open so it is the temperature for which $f_{DS} = 0.5$.

Figure 2.7 summarizes the data analysis described in this section. The figure shows a model example of $\Delta c(T)$, the fit with a Gaussian function, the straight-line baseline and the

calculated fraction of open base pairs.

All the calorimetric data presented in the result chapters of this work are obtained by measuring three identical samples and averaging the results. The error bars shown are the standard deviation of the three measurements.

2.3 Introduction to scattering theory

The theoretical and experimental descriptions in this section will cover only the techniques used for this work. Thus, no inelastic scattering is mentioned, neither the spin or magnetic contributions to the scattering are presented. Information for this section was collected from the books by Sivia [32] and Squires [33].

2.3.1 Basic principles of a scattering experiment

In a scattering experiment most of the information about the system is extracted from the change of momentum and energy of the scattered particles with respect to the incident particles. The momentum change is described mathematically by:

$$\vec{Q} = \vec{k_i} - \vec{k_f} \tag{2.7}$$

where $\vec{Q}$ is the momentum transfer (also called scattering vector), $\vec{k_i}$ and $\vec{k_f}$ are the initial and final wavevectors of the scattered particle. The work described in this document is based in elastic scattering in which $E_i = E_f$, so $k_i = k_f = \frac{2\pi}{\lambda}$, where λ is the wavelength of the incident particles. The vector diagram for an elastic scattering event is shown in fig. 2.8 where an incoming particle is deflected through an angle of 2θ. Trigonometry then leads to the result:

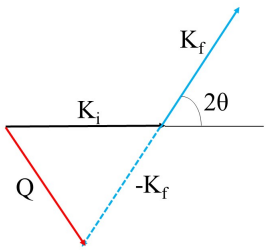


Figure 2.8: The vector diagram for elastic scattering through and angle of 2θ.

$$Q = \frac{4\pi \sin(\theta)}{\lambda} \tag{2.8}$$

An ideal elastic scattering experiment consists of determining the rate of incident particles which are scattered with a given momentum transfer. This can be encoded in a sample-dependent three dimensional function, $S(\vec{Q})$, called the scattering law.

Scattering cross section

The scattering cross section is used to quantify the scattering process. Assume a beam of particles of incident energy E_i is interacting with the target and a detector is set up in order to measure the scattered particles in a given direction with a given solid angle, $\Delta\Omega$, subtended by the detector with respect to the target. Then the differential cross section is defined by rate of particles arriving to the detector, R_{par}, normalized by the incident flux of particles per area per second (Φ_{fp}), by the solid angle $\Delta\Omega$ and by the number of scattering units in the target, N_s, which is normally the number of atoms or molecules in the target. The differential cross section can thus be expressed as:

$$\frac{d\sigma}{d\Omega} = \frac{R_{par}}{N_s \Phi_{fp} d\Omega} \tag{2.9}$$

Therefore the total scattering cross section or the number of particles scattered into all directions divided by the incident flux is:

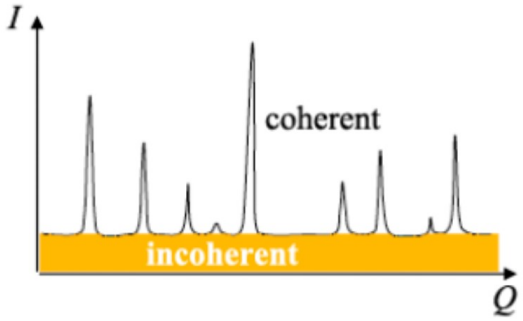


Figure 2.9: Example of coherent and incoherent contribution to the signal. Reproduced from [25].

$$\sigma_{tot} = \int_{alldir} (\frac{d\sigma}{d\Omega}) d\Omega = \frac{1}{N_s \Phi_{fp}} \int_{alldir} R_{par} \tag{2.10}$$

The total scattering cross section may be divided in two contributions, incoherent and coherent scattering: $\sigma_{tot} = \sigma_{coh} + \sigma_{incoh}$. Coherent signals are the result of pair correlation between the atoms. They give information about the probability of an atom being at a given position with respect to the rest of the sample. Bragg diffraction is an example of elastic coherent diffraction. Incoherent diffraction is due to the correlation of the atom with itself. It shows the variance from the mean structure of the sample and it is isotropic.

Practically, for our study, incoherent scattering appears as a Q-independent background in the total scattering signal (fig. 2.9) and coherent scattering is used to access the information of interest. It is important to note that a material with high incoherent scattering cross section, such as Vanadium, is useful for calibration of the efficiency of the detector in a neutron scattering experiment.

2.3.2 Scattering by a single fixed atom

Let us consider the simple event in which a beam of neutrons or X-rays is scattered by a single fixed atom. A stream of particles traveling in the z direction with wavelength λ can be described by a complex plane wave: $\Psi_i = exp(ikz)$. The result of the interaction of a planar wave with an atom is a spherical wave for both neutrons and X-rays assuming the X-rays are not polarized. In the case of neutrons the spherical symmetry arises from the fact that λ, of the order of 1 Å for thermal neutrons, is large in comparison with the size of the nucleus ($\approx= 10^{-5}$ Å). The theory of partial waves proves that under this circumstance the scattering comes entirely from the S-waves ($l = 0$) which are spherically symmetric [33]. For X-rays the interaction with an atom occurs with its electrons. The X-ray's electric field makes the electrons vibrate in the direction of the polarization of the incident wave. If the X-rays are not polarized the electrons vibrate isotropically and emit spherical waves. Hence, in this case, the scattering wave at a distance r from the atom takes the general form:

$$\psi_f = f(\lambda, \theta)\frac{e^{ikr}}{r} \tag{2.11}$$

where, 2θ is the scattering angle, k is the wavenumber (magnitude of the wavevector) and $f(\lambda, \theta)$ is the scattering length, which defines the probability of a particle of a given wavelength being deflected in a certain direction. This function differs greatly for neutrons and X-rays.

For neutrons $f(\lambda, \theta)$ tends to be invariant with respect to λ (for medium energy neutrons) and θ: $f(\lambda, \theta) = -b$. The magnitude of b determines the probability of the scattering and its sign indicates if the incident and outgoing waves change phase by 180°. The origin and mechanism of the scattering interactions quantified by the constant b are not totally understood but they depend on the nature of the nucleus and the direction of its spin relative to that of the neutron. Two important properties are that b is isotope-specific and it does not change with atomic number in a monotonic way.

In the case of X-rays, assuming that the energy of the incident beam is dar from the

resonance energy, $f(\lambda, \theta)$ decreases monotonically when Q increases. It has the same sign for all the elements and its magnitude is proportional to the atomic number Z. Thus, for X-rays, $f(\lambda, \theta)$ can be expressed as: $f(\lambda, \theta) = Zg(Q)r_e$; where Z is the atomic number, r_e is the classical radius of the electron, and $g(Q)$ is the atomic form factor. $g(Q)$ decays from 1 at the origin to 0 at large Q, its specific shape depends on the shape of the electron orbitals of the atom.

2.3.3 Scattering from an assembly of atoms

The scattering of neutrons by a conglomerate of N identical atoms will be presented next. Let the incident beam be described by a complex plane wave ψ_i. A given atom labeled as atom j at position R_j will make a contribution to the scattered wave at $\vec{r}$ position, $[\delta\psi_f]_j$:

$$[\delta\psi_f]_j = e^{i\vec{k_i}\bullet\vec{R}_j} f(\lambda, \theta)_j \frac{e^{i\vec{k_f}\bullet(\vec{r}-\vec{R}_j)}}{|\vec{r}-\vec{R}_j|} \tag{2.12}$$

$\vec{R}_j$ is the location of the j atom and $\vec{k_f}$ is the outgoing wavevector. The geometry of the event can be seen in fig. 2.10. If the superposition principle applies the total scattered wave can be calculated by adding up the contributions from the N atoms. The superposition principle applies if incident and scattered wave are considered plane waves and the interaction between the incoming particles and the target is weak, this two assumptions are known as the Born approximation [33].

Therefore, within the Born approximation the total scattered wave can be calculated as ψ_f:

$$\psi_f = e^{i\vec{k_f}\bullet\vec{r}} \sum_{j=1}^{N} f(\lambda, \theta)_j \frac{e^{i\vec{Q}\bullet\vec{R}_j}}{|\vec{r}-\vec{R}_j|} \tag{2.13}$$

where $\vec{Q} = \vec{k_i} - \vec{k_f}$ is the momentum transfer. Since the distance to the detectors where the particles are measured is usually much larger than the size of the samples a good

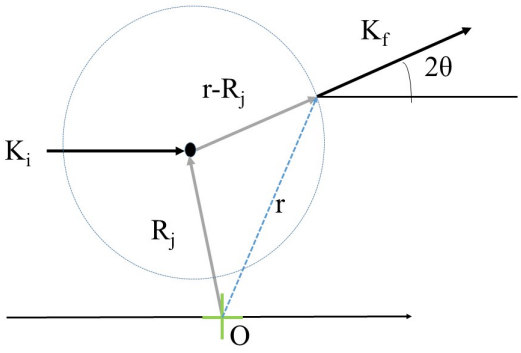


Figure 2.10: The contribution to the scattered wave from an atom j located at $\vec{R_j}$ relative to an arbitrary defined origin.

approximation is $|\vec{r} - \vec{R_j}| = |\vec{r}| = r$ so the denominator of eq. 2.13 can be replace with r and taken out of the summation. The rate of particles of wavelenght λ deflected elastically into a detector of differential area $dA = r^2 d\Omega$ is $R = |\psi_f|^2 dA = r^2 |\psi_f|^2 d\Omega$. Therefore the differential cross section (as defined in 2.3.1) can be related to the structure of the sample with an atom of type $f(\lambda, \theta)_j$ in the position R_j for $j = 1, 2, ..., N$ through:

$$(\frac{d\sigma}{d\Omega})_{el} = \frac{1}{N} |\sum_{j=1}^{N} f(\lambda, \theta)_j e^{i\vec{Q} \bullet \vec{R_j}}|^2 \tag{2.14}$$

And $A(\vec{Q}) = \sum_{j=1}^{N} f(\lambda, \theta)_j e^{i\vec{Q} \bullet \vec{R_j}}$ can be addressed as the complex scattering amplitude. Note that eq. 2.14 is the squared Fourier transform of the structure of the sample.

2.3.4 Fiber Diffraction

The importance of scattering data from DNA fiber samples for the present work justifies an introduction to fiber diffraction and how to interprete the classic scattering patterns of fiber DNA. This section is heavily inspired by the review of Lucas and Lambin [34]

One of the most important structural concepts for the interpretation of scattering data

of fibers is that of layer-planes and layer-lines in reciprocal space [35]. An easy way of introducing the concept is studying the scattering cross section of a simple polymer model.

Lets assume a single linear polymer consistent of identical point atoms (point scatterers) placed at equidistant positions $z_j = jp$ on the z-axis, where j is an integer and p is the polymer repeat period. The complex scattering amplitude of the previous section for such a system can be computed as [34]:

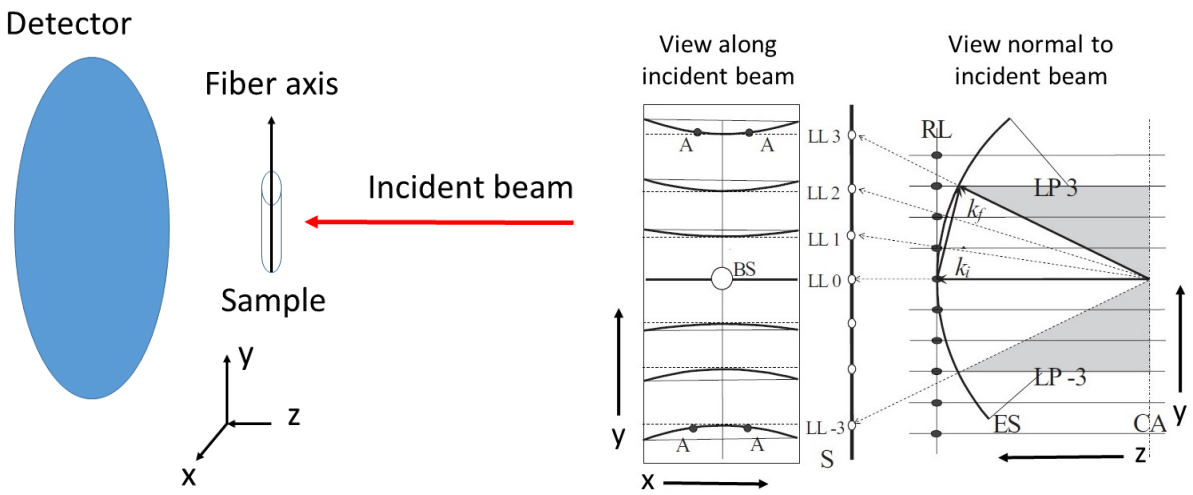
$$A(\vec{Q}) = g(Q)\frac{2\pi}{p}\sum_{l=-\infty}^{+\infty}\delta(Q_z - l\frac{2\pi}{p}) \tag{2.15}$$

where Q_z is the component of $\vec{Q}$ parallel to the z-axis (also parallel to the axis of the polymer). Therefore the Q_z for which the scattering amplitude is different from zero are quantized to integer multiples of $2\pi/p$. In reciprocal space, the planes in which $Q_z = 2\pi l/p$ are called layer-planes.

Note that for elastic scattering the values of k_f for which the diffracted intensity will be different from zero lie on the intersections of the layer-planes with the Ewald sphere [36]. This intersections are lattitudes on the sphere and constitute the bases of cones with common apex at the sphere's centre (as can be seen in fig. 2.11b reproduced from [34]). The intersections of the extrapolated conical surfaces and the plane of observation (S) give the position of the lines of non-zero diffracted intensity. These interceptions are called layer-lines.

When the incident radiation is perpendicular to the polymer, as in fig. 2.11 the diffraction pattern is symmetric with respect to the layer line 0 (usually called equatorial line). In many conditions, there is also symmetry with respect to the vertical line perpendicular to the equatorial line and any feature A will be reproduced identically in the four quadrants as seen in fig. 2.11b. If the incident beam is not perpendicular to the polymer axis the pattern is asymmetric.

The complexity of the model can be increased by assuming that the polymer is helical. In a helical polymer the monomer units making up the periodically repeated part of the



(a) (b)

Figure 2.11: a) Usual geometry of DNA fiber diffraction. b) Schematic of the construction of the layer-lines in the X-ray diffraction pattern of a linear array of point scatterers. RL stands for real lattice; ES for Ewald sphere; CA for cone axis; LP of layer-planes; S for observation screen; BS for beam stop and; LL for layer lines. Adapted from [34].

molecule are aligned along a regular circular helix. The full general theory for calculating the form factor of a helix was published by Cochran, Crick and Vand [37] and is summarized with the CCV formula.

This formula computes the scattering amplitude of a continuous helix of radius R and pitch P, for such a system the cylindrical coordinates $\vec{r_j} = (\rho_j, \varphi_j, z_j)$ of identical atoms placed at axial intervals p_a can be defined with relation to an arbitrary atom taken as the origin (R, φ_0, z_0). The step-by-step derivation can be found at [37] and the formula is:

$$A(\vec{Q}) = \frac{2\pi}{P} f(Q) \sum_{n,m} J_n(Q_r R) e^{in(Q_\varphi - \varphi_0 + pi/2)} e^{(Q_z z_0)} \delta(Q_z - n\frac{2\pi}{P} - m\frac{2\pi}{p_a}) \tag{2.16}$$

where $\vec{Q} = (Q_r, Q_\varphi, Q_z)$ in cylindrical coordinates and $J_n(x)$ is the cylindrical Bessel functions of integer order n.

The δ-function in eq. 2.16 makes it clear that the diffraction patterns will still be organized in layer lines. The δ-function imposes the selection rule:

$$\frac{l}{P_a} = \frac{n}{P} + \frac{m}{p_a} \tag{2.17}$$

With this selection rule the δ-function can be rewritten and it can be concluded that non-zero intensity will be seen at $Q_z = 2\pi l/P_a$.

Thus the scattering pattern of a phosphorous helix in a single strand DNA backbone can be calculated with eq. 2.16 setting $P_a = P = 10p_a = 34$ Å^{-1}and $R = 10$ Å^{-1}. The result can be seen in fig. 2.12 reproduced from [34]. The figure shows a very characteristic cross pattern.

The selection rule for this system is then $l = n + 10m$. As an, example the selection rule gives $(n = 0, m = 0)$ (i.e. J_0) as leading term in the diffraction amplitude for the equatorial line ($l = 0$). For the $l = 1, 2, 3, 4, 5$ the equation selects as leading terms $(1, 0)$, $(2, 0)$, $(3, 0)$, $(4, 0)$, $(5, 0)$. Keeping in mind that n is the order of the Bessel function the cross pattern of the first five lines can be explained. Moreover the same selection rule imposes the repeat of the cross patterns along the direction of the axis of the helix every ten layer-lines. This creates a diamond pattern in the diffraction picture with the interior of the diamond approximately empty of intensity. These diamonds are characteristic of an helical distribution of matter in the diffracting object.

In order to complement this theoretical introduction to fiber diffraction, two very important results will be commented. They will allow the main differences between the secondary A and B form of DNA to be highlighted, which will be significantly important for the discussion of future chapters.

Fig. 2.13 show diffraction diagrams of Na-DNA fibers. The measurements were made with a geometry similar to that shown in fig. 2.11a. The fiber comprises a macroscopic number of long, parallel DNA molecules, Na^+ ions and water molecules. The cations and

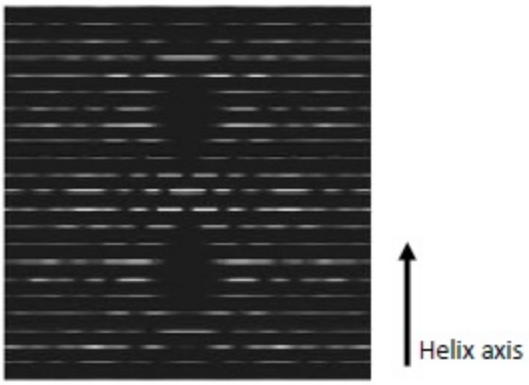


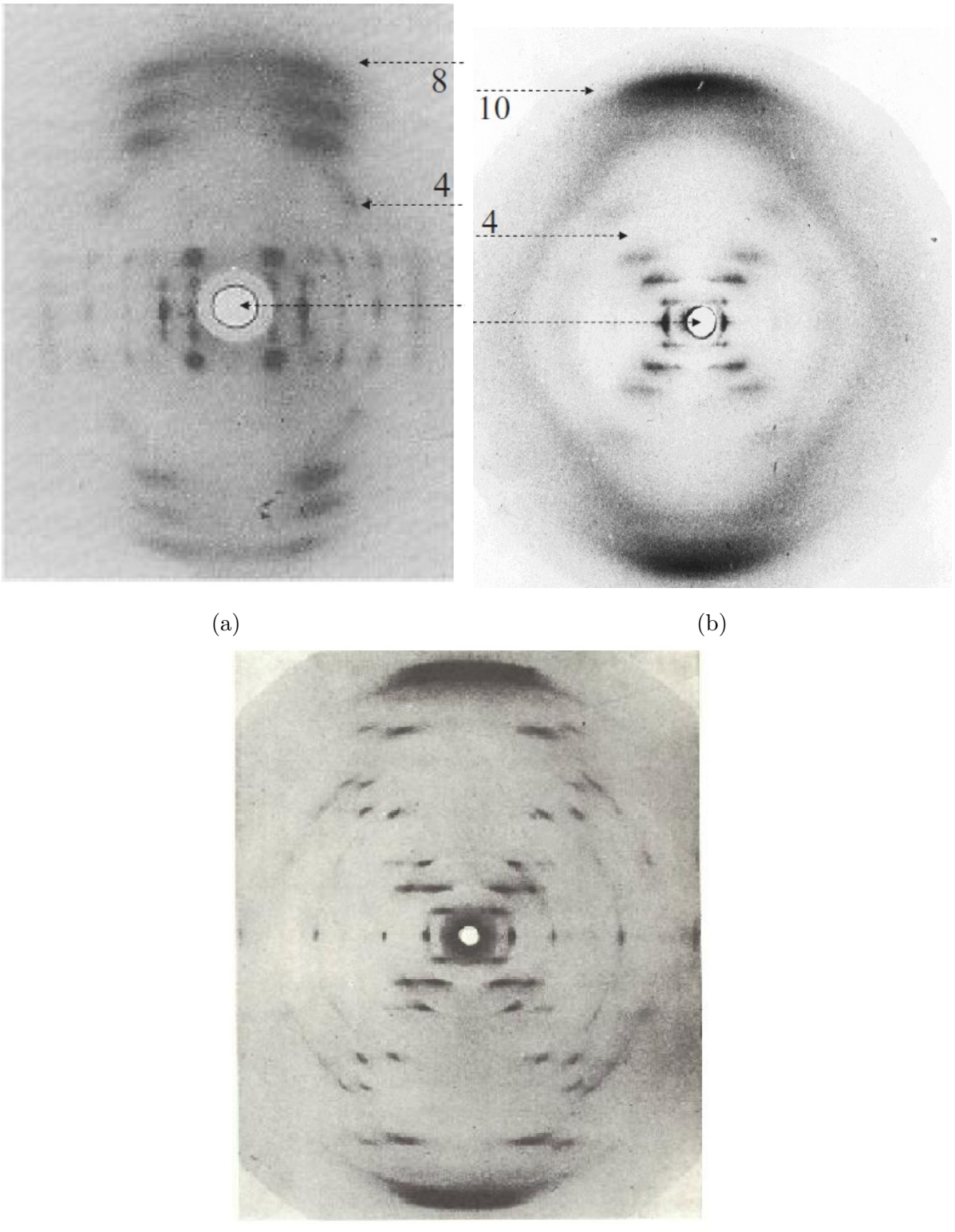
Figure 2.12: Simulated diffraction pattern by an atomic phosphorus helix. Reproduced from [34].

the water are in amorphous state and give only diffuse scattering. Fig. 2.13a shows the pattern classically associated with the A-form of DNA, Fig. 2.13b the one associated with the B-form.

After the discovery of these two forms of DNA it was proved that the transition between A to B form is reversible [39, 40]. In these works the water content of a sample was modified between two states. The drier state of the sample gave pattern A and was interpreted as crystalline, with long range order, due to the sharp spots specially close to the center. The wetter state of the sample gave pattern B and due to its broader features was called semycristalline and was considered to have less long range order due to the disordering effect of the increased amount of water surrounding the molecules.

Both patterns are organized in regularly spaced layer-lines as expected from long molecules having a repeating structural unit (the base pair). The dramatic change of the features between patterns was interpreted not only as a change in the long range arrangement or the sample but as a change in the internal structure of the DNA molecules. This is why the DNA is said to adopt different conformations depending on the environmental conditions. The discovery of A and B form was followed by many other conformations such as C, D, Z...[41].

The B-form is believed to be the conformation of chromosomic DNA *in vivo* and will be



(a) (b)

(c)

Figure 2.13: X-ray fiber diffraction patterns of a) A-DNA, b) Semycristalline B-DNA. Reproduced from [34] and c) Crystalline B-DNA reproduced from [38].

the focus of the present work. Its pattern is qualitatively very similar to fig. 2.13, showing the helical structure of the molecule. The reflexion of the 10^{th} layer line of this form will be instrumental in the study of the melting transition in DNA fibres.

Soon after the discovery of the B-form it was proven that this reflexion was related with a nucleotide repeat of 3.4 Å^{-1} [42]. This reflection was modeled as the result of a concatenation of flat nucleotides (even though in reality the sugar and nitrogenous base do not lie in the same plane) stacked on top of each other and covalently linked to the phosphate backbone. The bases scatter coherently despite being irregular in their chemical composition and geometrical structure because they behave like thin scattering slabs seen edge-on by the incident beam (perpendicular to the fiber). In summary the stacked base pairs act somewhat like a parallel slit grating to produce the large 10^{th} layer reflections.

Since the B-pattern shows ten layer-intervals separating the center from this reflexion it follows that in this conformation the pitch of the DNA is 34 Å^{-1}. The cross pattern in the center of the picture is the second strongest feature, it is, as stated above, a footprint of the helical conformation of the DNA backbone. A last crucial piece of information that can be extracted easily from the B-pattern is the radius of this form ($\approx$ 10 Å^{-1}), which is patent due to the distance between the equatorial reflections.

Fig. 2.13c shows a pattern very similar to fig. 2.13b but the blurred reflections are substituted with sharp points and lines which indicates both samples are in the same conformation but the former has better crystalline order. Because of this, patterns like fig. 2.13b are usually said to relate to the "semicrystalline" B-form and patterns like fig. 2.13c are related to the crystalline B-form.

The A-form has 11 nucleotides per pitch with an average distance between base pairs of 2.56 Å. The removal of water with respect to the B-form contracts the molecule along its axis and, the base pairs become inclined with respect to the horizontal plane (with a tilt angle as large as 20°). The effect of this tilt is that most of the base pairs are not seen edge-on by the incident beam and thus they no longer act as a horizontal slit grating. The base pair stacking no longer gives a strong reflexion on the 10^{th} layer but two pairs of smear

reflexions appear on the sixth, seventh and eighth layer lines which are sometimes refereed to as off-axis Bragg peaks.

2.3.5 Small angle scattering

When studying the scattering from a conglomerate of atoms, the characteristics will depend on the spacial distribution, number and $f(\lambda, \theta)$ parameters of the atoms. The spacial distribution weighted by the $f(\lambda, \theta)$ of each atom can be quantified by the so called scattering length density function, β. The scattering length density at a position $\vec{r}$ in a system is given by: $\beta(\vec{r}) = \rho(\vec{r})f(\bar{\lambda}, \theta)(\vec{r})$ where $\rho(\vec{r})$ is the local atomic number density and $f(\bar{\lambda}, \theta)(\vec{r})$ is the mean scattering length which varies as a function of atom and isotope for neutrons and as the Thomson scattering length times the local number of electrons for X-rays [43]. Then a continuum generalization of the discrete case of eq. 2.14 can be made:

$$(\frac{d\sigma}{d\Omega})_{el} \propto \frac{N}{V}\langle|\int\int\int_V \beta(\vec{R})e^{i\vec{Q}\bullet\vec{R}}d^3\vec{R}|^2\rangle_\Theta \tag{2.18}$$

where the system contains N particles in a volume of V and the angular brackets indicate an ensemble average over all the orientations of the particles.

As mentioned previously (sec. 2.3.3), the differential cross section equates the Fourier transform of the scattering length density of the sample. A basic property of Fourier transforms is the inverse relationship between length scales in real, $\vec{r}$, and reciprocal space, $\vec{Q}$. Thus experiments focusing in small Q ranges provide low resolution structural information but are optimal for studying the configuration of macromolecules in the mesoscopic scale (such as short DNA chains). Equation 2.8 demonstrates that low-Q experiments should prioritize measurements at small θs and use large λs. Small diffraction angles are reached by having a long flight-path between sample and detector.

Many small angle scattering experiments are performed on solutions so the sample consists of N object of interest placed within the solvent. The small angle scattering (SAS)

experiments described in the present work were performed on a dilute solution of short chain DNA molecules so the objects of interest were N isolated molecules isotropically oriented and randomly located within the sample. Under these circumstances equation 2.18 reduces to a one dimensional integral:

$$\frac{d\sigma}{d\Omega} \propto \int_0^{d_{max}} P(r)\frac{sin(Qr)}{Qr}dr \tag{2.19}$$

where d_{max} is the longest dimension of the scattering object. $P(r)$ is the pair-distribution function. It represents the average of the shapes of the scatterers in the solution. P(r) is usually calculated by the inverse Fourier transform of the cross section and it will be essential for the interpretation of the SAS data in the present work.

2.4 Diffraction Instruments

In this section an overview will be presented about the instruments used for the collection of data in this work.

2.4.1 Neutron diffractometers

IN3

IN3 is a three-axis spectrometer at the ILL (France), it was used for sample characterization. This technique provides access to the scattering function $S(Q, w)$ in a large volume of the reciprocal space. However, only elastic measurements were performed. Three-axis instruments have a mobile rotational axis at both the sample and analyzer positions. The direction of the initial and final wavevectors with respect to the sample is defined by the angles of each axis. Any combination of k_i and k_f can be selected as long as the vector triangle $\vec{k_i} - \vec{k_f} = \vec{Q}$ can close.

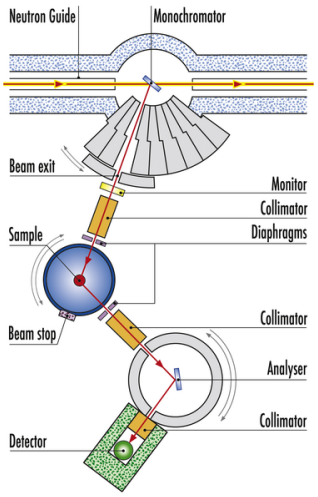


Figure 2.14: Scheme of the three axis spectrometer IN3 at the ILL. Reproduced from the ILL website.

For the experiments presented in this document a PG(002) monochromator was used delivering a $\lambda = 3.3545$ Å^{-1}, another PG(002) crystal was used as analyzer, $40'$ collimation was used before and after the sample position and a PG filter was placed before the sample to filter high order contamination. The measurements were performed in reciprocal Ångströms (Å^{-1}) with an orthogonal co-ordinate system.

D16

The diffractometer D16 at the ILL is used for small angle scattering, powder diffraction and single crystal studies. It uses a pyrolitic graphite monochromator, a beryllium filter, two sets of slits for incident beam collimation (one just after the monochromator and another just before the sample) and a square ^3He detector of area $320 \times 320\ mm^2$. Fig. 2.15 presents a schematic of the instrument. This instrument was used for the collection of reciprocal space maps to structurally characterize the high oriented DNA fibers in PEG and ethanol solutions at room temperature. The wavelength was set to 4.5 Å and the scattering from each sample was recorded in two different orientations: with the fiber axis parallel and perpendicular to the scattering plane. The first orientation allows to study periodic order in the sample along

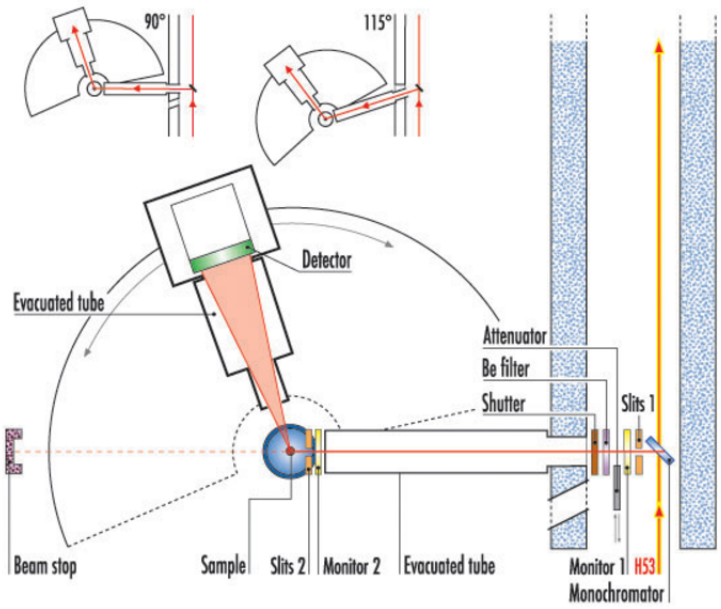


Figure 2.15: Scheme of the diffractometer D16 at the ILL. Reproduced from the ILL website.

the fiber axis and the latter orientation gives information about the lattice spacing of the sample in the direction perpendicular to the fiber axis. A cylindrical piece of vanadium, of length similar to the sample, was measured for every detector position to account for the efficiency of the detector.

D19

D19 is a thermal diffractometer at the ILL, specially designed for studying large structures and fibers. D19 was configured with a graphite monochromator delivering an incident wavelength of 2.4 Å. The monochromator-sample distance was 3.18 m. Pre-sample collimation and beam size were defined with two sets of squared slits ($12 \times 12\ mm^2$) and a circular aperture (diameter of 10 mm). The slits were at 1.51 m and 2.49 m from the monochromator. The circular aperture was at 3.1 m from the monochromator. High-order Bragg contamination was eliminated by using a graphite filter. D19 has a 120°position-sensitive detector which was fixed throughout the experiment. There was no collimation between sample and detector. Fig. 2.16 presents a scheme of the instrument. A cylindrical piece of vanadium,

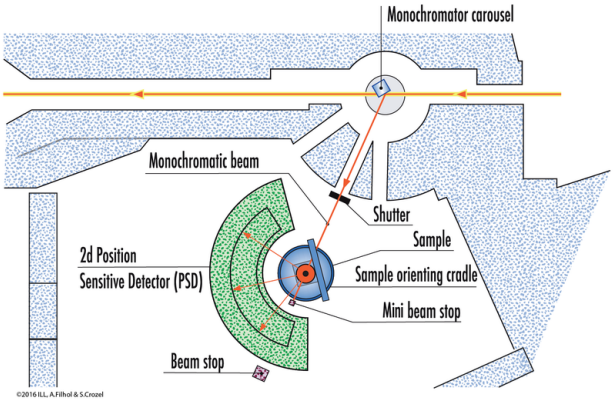


Figure 2.16: Scheme of the D19 diffractometer at the ILL. Reproduced from the ILL website.

of length similar to the sample, was measured for every detector position to account for the efficiency of the detector.

This instrument was used to study the diffraction pattern of samples of fiber DNA submerged in PEG solutions and ethanol-d6/^2H_2O mixtures as a function of temperature through the melting transition. Reciprocal space maps with the fiber axis within the scattering plane were collected at several temperatures between 20 °C and 110 °C.

The stabilization time between measurements when the temperature was changed was 15 *min*. On this instrument two different sample environments were used:

a) For the fibers submerged in PEG 6000 at 17% (w/w) two heaters were attached to opposite vertices of the square cassette. At room temperature the temperature stability was around 0.2 °C but thermal oscillations increased with temperature. Close to the melting temperature of this sample ($\approx$ 95 °C) the stability was around $\pm$1 °C.

b) For the ethanol samples the sample environment was a cylindrical aluminum chamber with a flow of thermalized air. Stability of the sample temperature during the measurement was $<$ 0.1 °C over the entire temperature range.

WOMBAT

Wombat is a high intensity powder diffractometer at the OPAL research reactor (Australia). It was used with a pyrolitic graphite monochromator (which gives access to a range in λ from 2.4 to 5.8 Å), a Be filter for avoiding high-order Bragg peak contamination and a radial oscillating collimator for background suppression. The detector is a PSD of 120°$\times$ 200 mm height. A cylindrical piece of vanadium, of length similar to the sample, was measured for every detector position to account for the efficiency of the detector.

Only fibers submerged in PEG solutions were measured on this instrument. Two different set of measurements were performed: Firstly λ was set to 5.65 Å and room temperature data were collected to generate wide reciprocal space maps for both longitudinal and transversal orientations for each sample as done in D16. Secondly, λ was changed to 4.62 Å and the temperature dependence of the diffraction patterns were studied for each sample. This is analogous to what was done in D19.

The design and performance of the sample environment used in WOMBAT is similar to the aluminum chamber with thermalized air-flow used on D19. Due to a technical problem, associated with maintaining a constant gas flow, during the measurement of the fibers submerged in PEG 6000 20% (w/w) this sample suffered of thermal oscillations of around 1 °C at the highest temperatures. For the other samples measured at WOMBAT the thermal stability was < 0.1 °C over the entire temperature range.

D22

D22 is a small angle neutron diffractometer at ILL which is used mainly for the study of nanoscale samples. Fig. 2.17 shows a scheme of the instrument. A velocity selector gives a monochromatic beam (with $\Delta\lambda/\lambda = 10\%$) with a λ range from 4.5 to 40 Å. Eight collimation sections adjust the virtual source-sample distance. The samples were placed in an aluminum rack with an XYZ position controller. The ^3He based multi detector has a pixel size of 0.8 $\times$ 0.8 cm^2 and can move over a distance of 1-20 m in a vaccum tube.

D22 was used to collect SANS patterns as a function of temperature from solutions of short chain DNA containing the Widom-601 sequence (sec. 2.1.2). The solution was loaded in a rectangular crystal cuvette in which the path-length through the sample is of 2 mm for the neutrons and 0.5 mm for the UV-Vis beam (Fig. 2.18). A reference cuvette containing only buffer (no DNA) was also prepared and measured at several temperatures. Both sample and reference were degassed in a vacuum oven for three hours before the experiment started. In-situ UV-Vis spectroscopy was recorded with the neutron data for each temperature. The UV-Vis set up was purchased from IDIL company. It consisted of a UV-Vis spectrometer (HR2000) able to measure absorbance and fluorescence in a range of wavelengths between 200 and 950 nm connected to two bifurcated optical fibers which allowed to measure alternatively the absorbance of the sample and the reference buffer. Fig. 2.19 shows an schematic of the setup of the UV-Vis device. Neutrons of wavelength 6 Å were used to record the SANS pattern from the sample and reference at several temperatures between 20 and 95 °C. UV-Vis transmission was recorded at each temperature and also during stabilization time. A thermalized water-bath was used for heating and cooling and the temperature of the sample was measured with a thermocouple inserted in the aluminum rack. Scattering data were recorded at two sample-detector distances for each temperature: 5 and 17 m, with collimation of 5.6 and 17.6 m respectively. Merging these two sets of data gave a Q-range of $0.003 < Q < 0.15$ Å.

2.4.2 X-ray devices

X-rays can be generated in lab sources where photons are created by de-excitations of electrons or in synchrotron facilities in which electrons with a circular trajectory emit radiation of wavelengths related to their acceleration. In the present work only lab sources were used. The instrument used were a Copper-based X-ray diffractometer and a small angle X-ray scattering (SAXS) instrument.

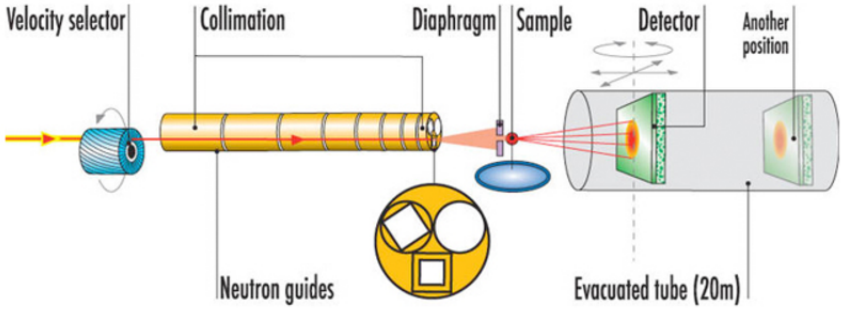


Figure 2.17: Scheme of the D22 SANS instrument at the ILL. Reproduced from the ILL website.

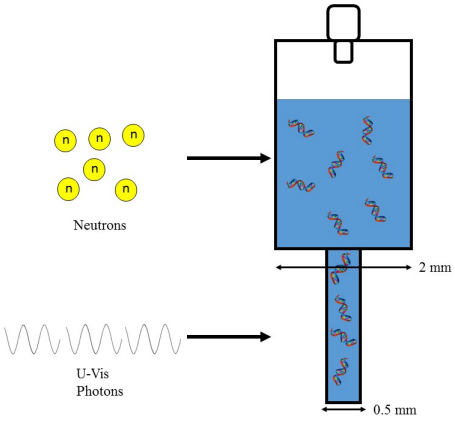


Figure 2.18: Diagram of the crystal cell used for the SANS experiment.

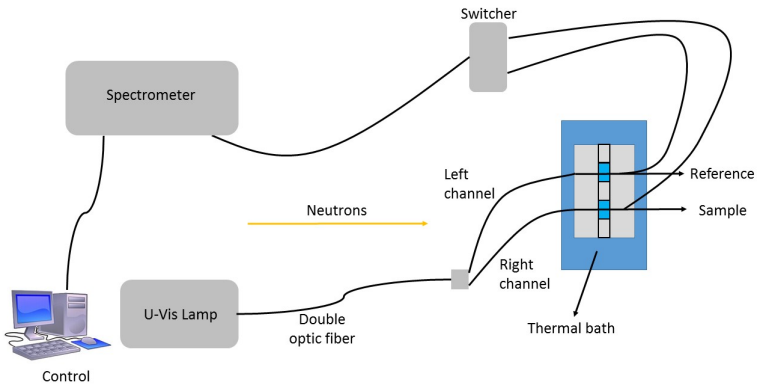


Figure 2.19: Schematic of the set up of the in situ UV-Vis absorption device. The Switcher selects between the channels so sample and reference can be measured alternatively. The spectrometer is controlled by the same computer that runs the neutron measurements so absorption scans and collection of scattering data can be easily automatized.

GeniX source

An X-ray lab device at the EMBL at Grenoble (instrument responsible: Hassan Belrhari) was used to collect wide angle diffraction patterns of DNA fibers. It is the combination of a GeniX X-ray generator from Xenocs company and a Mar345 fast image plate detector.

In the lab source used for the wide diffraction studies, X-ray were generated by means of a vacuum tube with a cathode and an anode. Electrodes collide with the anode and generate X-rays via a resonance effect which produces the so called CuKα radiation with $\lambda = 1.54$ Å. The Mar345 is a 345 mm-diameter image plate. Image plate detector are based on photostimulated luminescence, which is the release of energy within a phosphor by radiation with visible light. So in order to obtain an image one should illuminate the detector plate twice. First with the X rays diffracted from the sample, which "write" the image. Then the detector is illuminated with a laser which "reads" the plate. Photons in the visible range of the spectrum are released during the reading and they are measured with a phosphoimager for generating the final diffraction pattern.

The experimental protocol for this instrument was as follows: Small pieces of the DNA films were placed in quartz capillaries of thickness 0.01 mm and outside diameter 2 mm with the fiber and the capillary axis parallel. For humidified samples, the fibers inside the capillary were left to equilibrate with a humid atmosphere for 3 days before being sealed with wax. For the samples submerged in PEG solution or ethanol mixtures, a very small volume of the liquid was degassed and introduced in the capillaries with a small syringe making sure the DNA was properly soaked; these samples were equally sealed with wax. The samples were placed in the X ray device with the fiber axis roughly perpendicular to the incoming beam and the normal to the film parallel to the incoming beam. The sample-detector distance was 175 mm for most measurements and the counting time was 2 hours per pattern.

SAXS device

The VAXTER instrument at the Friedrich-Alexander University of Erlangen-Nuremberg (Germany) was used for the collection of SAXS data. Fig. 2.20 is a photograph of the apparatus. It was originally developed by SAXLab in Denmark and customized for applications of interest and optimal efficiency by the Condensed matter physics group at the universistu. It includes a 70 kV GaMetalJet D2 source which provides a Ga-K$\alpha_{1,2}$ beam (wavelength of 1.34 Å^{-1}). Collimation is provided by two sets of double slits which limit beam size to 0.4×0.4 mm^2 at the sample position. The temperature of the sample was controlled by a water-circulation system. The detector was set at 3.5 m from the sample. The range of the data recorded was $0.006 \leq Q \leq 0.2$ Å^{-1}. The whole X-ray path of the instrument was evacuated.

2.5 Data collection and reduction

The procedures used to reduce the scattering raw data to the final state presented in the result chapters will be described here in detail.

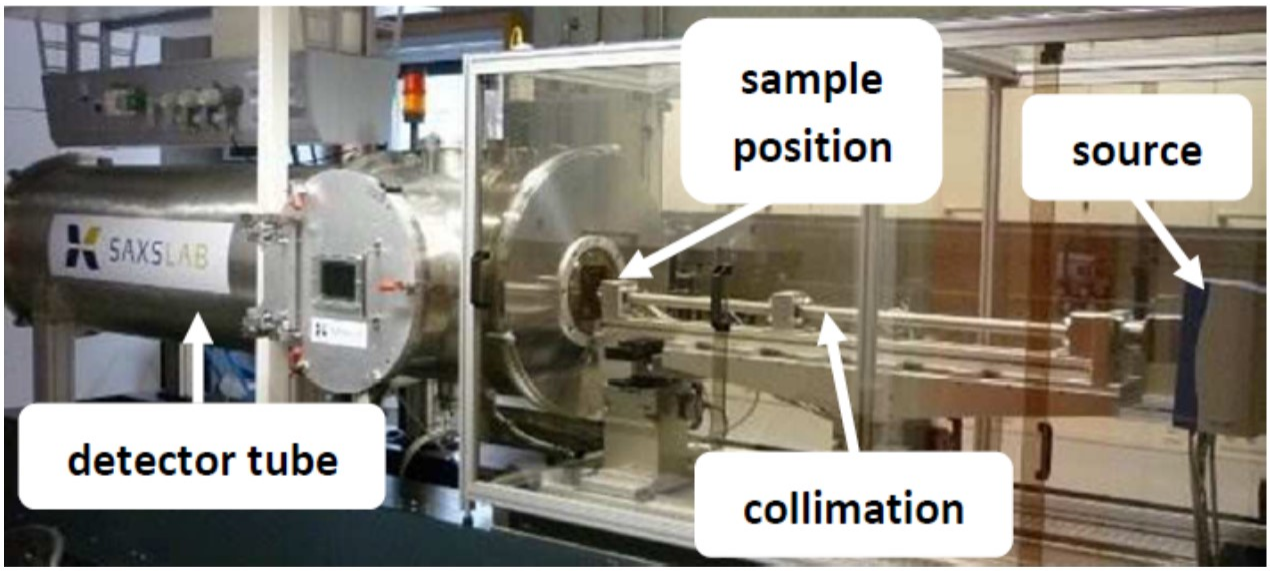


Figure 2.20: Photograph of the SAXS machine used in this work. Reproduce from the supporting info of [44].

2.5.1 Wide angle neutron scattering

Although three different diffractometers (D16, D19 and WOMBAT) were used in the experimental program on the present work, the data collection and reduction were fundamentally identical in each case. The scattering of each sample at each orientation and temperature was measured in transmission through the sample cassette with an incident beam defined by $\vec{k_i}$ scattered through a 2θ angle into a final direction with wavector k_f, thus the momentum transfer vector is $\vec{Q} = \vec{k_i} - \vec{k_f}$.

A scan was performed for each detector position (in the case of D19 and WOMBAT there was only one detector position, that is not the case for D16), the sample was rotated around the axis normal to the scattering plane. The rotation of the sample was quantified by ω which is the angle between the normal to the sample cassette and the direction of the incident beam (fig. 2.21a). The raw data consist of a detector image with the scattered intensity as a function of 2θ and detector height for every ω value.

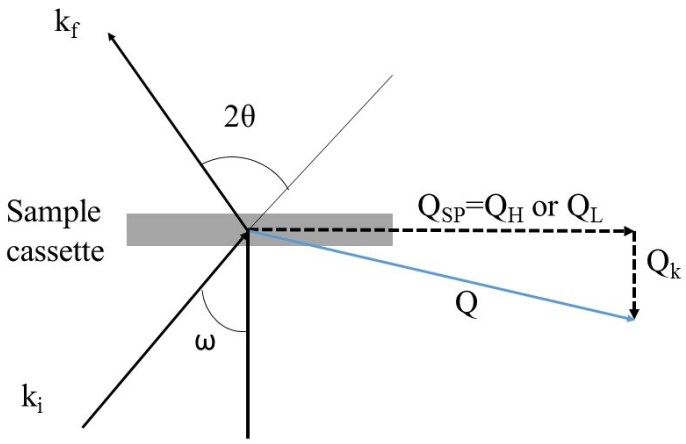
The first step in the analysis was to reduce the raw data to reciprocal space maps. It was assumed that the most convenient way of presenting such maps was by defining a system of reciprocal coordinates with reference to the sample cassette. Such a reference system has been used in previous articles ([31],[45]). As presented in fig. 2.21 the momentum transfer vector $\vec{Q}$ can be expressed as $\vec{Q} = (\vec{Q_H} + \vec{Q_K} + \vec{Q_L})$ where $\vec{Q_H}$ is parallel to the axis of the DNA fibers, $\vec{Q_K}$ is perpendicular to the axis and to the face of the cassette, lastly $\vec{Q_L}$ is orthogonal to the previous two.

In order to present the raw data in terms of Q_H, Q_K and Q_L each detector image was first normalized to monitor and divided by the vanadium data corresponding to the same detector position (detector efficiency correction). Integrating over the height of the detector gave a map of intensity as a function of ω and 2θ for each detector position, an example can be seen in fig. 2.22a. In the case of D16 the different maps for each detector position were assembled together and in the parts in which two maps overlapped in the 2θ-ω space an average was taken as the representative scattered intensity (fig. 2.22b).

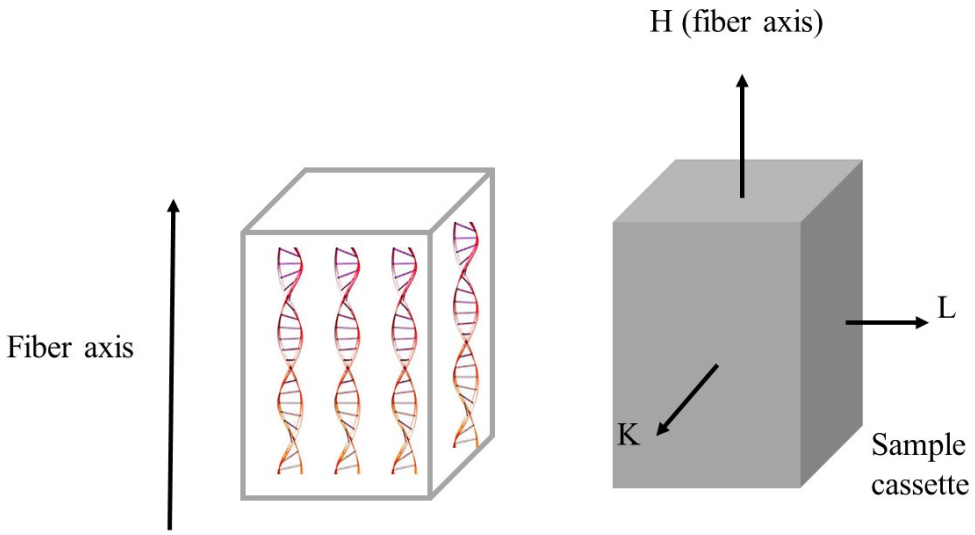
The "ω-2θ" maps in fig. 2.22 present a clearly not natural dip in the intensity whose position evolves linearly with 2θ. It is an artifact due to the self attenuation of the sample when the scattering angle of the deflected neutrons is so that their direction is in the plane of the cassette. The data must be corrected for this. In order to do so the transmission factor of the sample was calculated using:

$$\mathcal{T}(\omega, 2\theta) = \frac{1}{A_b} \int_x \int_y e^{-\mu t(x,y,\omega,2\theta)} dxdy \times \frac{1}{\cos\omega} \tag{2.20}$$

where $\mathcal{T}$ is the transmission factor, μ is the linear attenuation coefficient, A_b is the area of the incident beam and t is the path length of a given set of neutrons through the sample in function of ω, 2θ and (x,y) which is the point in the sample at which the scattering event occurred with reference to an arbitrary origin in one of the vertices of the sample (fig. 2.23). The facor $1/\cos\omega$ was included to account for the footprint of the beam on the sample as the sample ratated. Since the linear attenuation factor for the sample is not known, μ was firstly approximated by the linear attenuation factor of ^2H_2O and them adjusted slightly for

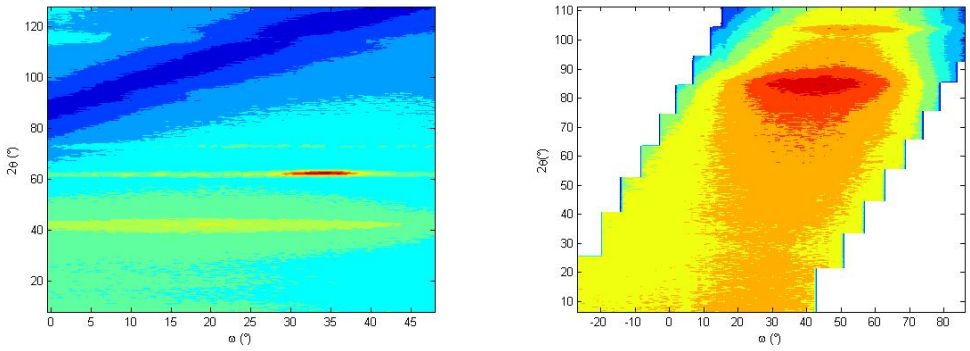


(a)



(b)

Figure 2.21: a) and b) Schematic of the reciprocal coordinate system defined with respect to the DNA fiber axis. $\vec{Q_H}$ is parallel to the fiber axis, $\vec{Q_K}$ is perpendicular to the fiber axis and to the face of the cassette, $\vec{Q_L}$ is orthogonal to both $\vec{Q_H}$ and $\vec{Q_K}$. $\vec{Q_{SP}}$ is in the scattering plane and it matches $\vec{Q_H}$ in the longitudinal orientation and $\vec{Q_L}$ in the traversal orientation.



(a) D19 (b) D16

Figure 2.22: Scattered intensity as a function of ω (sample rotation) and 2θ for Na-DNA submerged in a PEG 6000 solution at as measured at room temperature on a) D19 and b) D16.

the result to match the attenuation observed in the experimental data.

An example of part of the calculation will be presented here. The integral was divided into different cases so that in each case the trajectory of the neutrons, t, is defined as a single function of x, y, ω and 2θ. The simplest case assumes that $\theta_1 < 0$ and $\theta_1 < \theta_2 < \theta_{max}$ where $\theta_{max} = atan(\frac{w-B_w}{2d})$, as is highlighted in fig. 2.23. Under these assumptions every possible path of the scattered neutrons through the sample enters it by the face A, leaves the sample by face C and the path length can be described as:

$$t = \frac{d-y}{\cos\theta_1} + \frac{y}{\cos\theta_2} \tag{2.21}$$

Then $\mathcal{T}$ is computed with eq. 2.20 with the limits for the integration in x being $(w - B_w)/2 < x < (w + B_w)/2$ and in y being $0 < y < d$.

Fig. 2.24 shows an example of the result of the correction for both a single value of ω and the complete ω-2θ map.

The corrected intensity map was then transform to the Q_H-Q_K space for the data of the longitudinal orientation (fiber axis within scattering plane) or the Q_L-Q_K space for

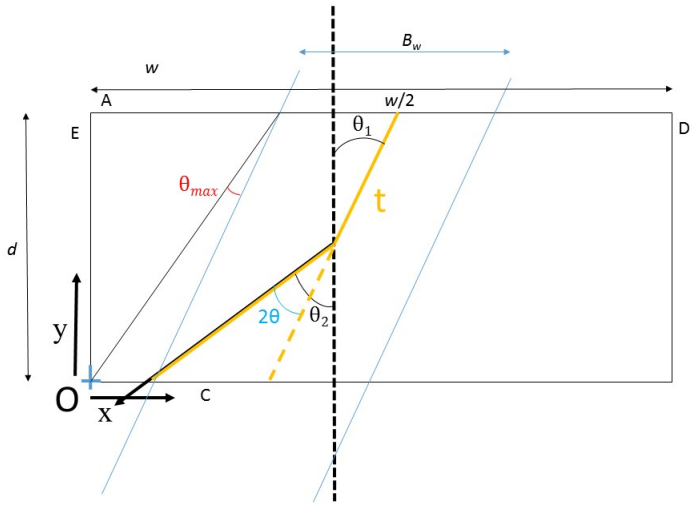
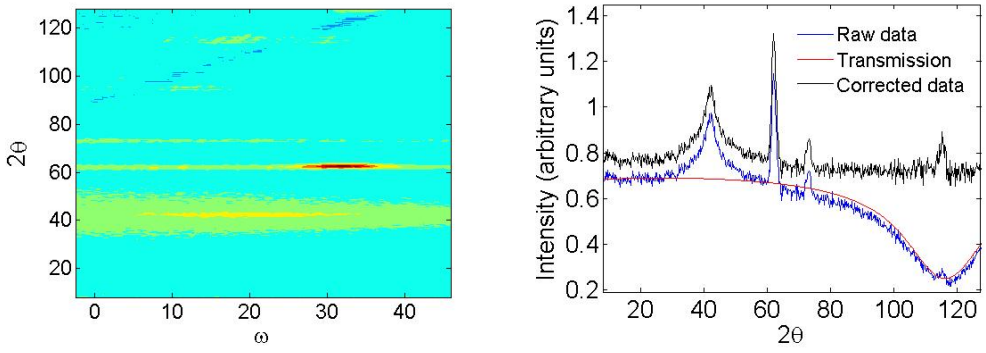


Figure 2.23: Geometry of the calculation of the transmission factor $\mathcal{T}$. $\theta_1 = \omega + \Delta\theta$ and $\Delta\theta$ is an instrumental offset. w is the width of the sample, d is the thickness of the sample. B_w is width of the neutron beam. The center of the neutron beam is supposed to go though the middle of the sample in the front face. θ_{max} is a limit for θ_2 in the example of the calculation of T and it is described in the text.



(a) (b)

Figure 2.24: a) Example of ω-2θ map after the self attenuation correction. b) Intensity vs. 2θ for the raw data (blue), self-attenuation corrected data (black) and value of the transmission factor, T', (red) for a given omega value.

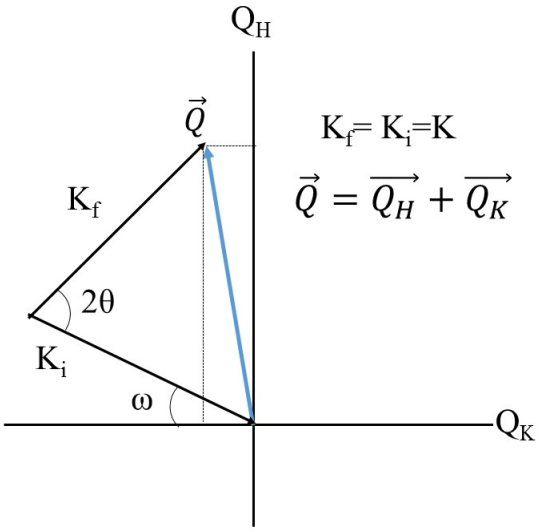


Figure 2.25: Diagram of the scattering geometry relating the coordinates ω and 2θ to the Q_H-Q_K space. In the case of the transversal orientation the diagram would be identical and Q_H would be replaced with Q_L.

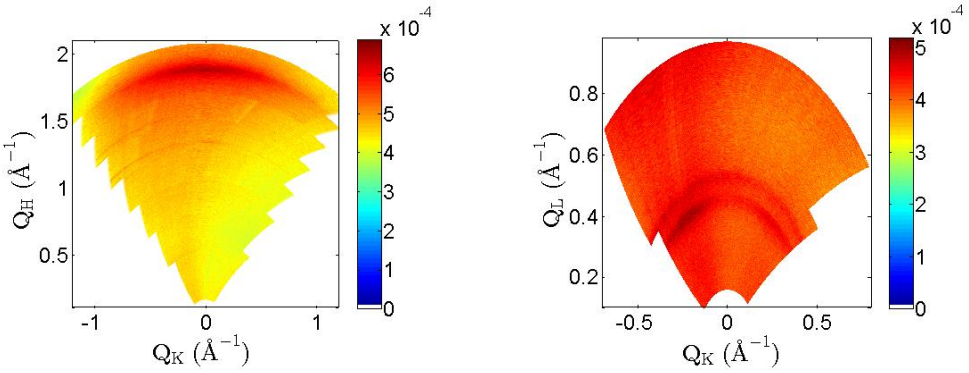
the transversal orientation (fiber axis normal to the scattering plane). This is done by transforming the ω-2θ space using the equations:

$$\begin{aligned} Q_{H,L} &= k(\sin\omega + \sin(2\theta - \omega)) \\ Q_K &= k(\cos\omega - \cos(2\theta - \omega)) \end{aligned} \tag{2.22}$$

where $k = 2\pi/\lambda$. These equivalences are apparent when following the scattering geometry of fig. 2.25.

The final reciprocal maps for both orientations can be seen in fig. 2.26. The slight drop in the intensity which is observable in the upper left corner of fig. 2.26a proves that the attenuation correction is not perfect. Since such marginal defect is relatively far away from the features of interested studied in this work it can be reasonably assumed that its effect in the interpretation of the data is negligible.

Before interpreting the data of the maps with the fiber axis normal to the scattering plane (transversal orientation) another reducing step was carried on. The data of these



(a) (b)

Figure 2.26: Reciprocal space map of a DNA sample submerged in PEG solution as measured in D16 for both orientations: fiber axis in the scattering plane (a) and normal to the scattering plane (b).

maps were summed over the sample rotation angles (ω) and plotted as a function of the 2θ of the magnitude of the scattering vector (Q) which are related through ec. 2.8).

For the reciprocal maps collected with the fiber axis parallel to the scattering plane it was important to extract a representative data of a scan along the helix axis, i.e. a scan along Q_H when $Q_K = 0$. To get this, a narrow Q_K-range was chosen centered at $Q_K = 0$ (usually $-0.04 < Q_K < 0.04$ Å^{-1}) Data whithin this range were extraxted and plotted against their Q_H values, disregarding the Q_K values.

2.5.2 Small angle neutron scattering

The raw SANS patterns were corrected for environmental background using a blocked beam measurement. Defects in the sample and reference cells were accounted for with empty cell measurements. Afterwards the intensity in absolute units (cm^{-1}) was obtained by using the formula which relates the experimental elastic signal with the differential cross section:

$$I_{el}(\vec{Q}) = I_0 \eta(\lambda) T_r(\lambda) \Delta\Omega (\frac{d\sigma}{d\Omega})_{el} \otimes R(\vec{Q}) \tag{2.23}$$

where I_0 is the incident flux, η is the efficiency of the detectors, T_r the transmission of the sample, $\Delta\Omega$ is the solid angle of the detector with respect to the sample and $R(\vec{Q})$ is the instrumental resolution. The experimentally detected intensity is smeared with respect to the real scattered intensity due to the finite size of the incoming beam, the wavelength resolution and the pixel size of the detector. $R(\vec{Q})$ describes the distribution of the Q vectors at a given instrument configuration and corrects for this smearing.

A sheet of plexiglass was measured in order to account for the detector efficiency (η) since it scatters isotropically, transmission of the sample was approximated by attenuating the beam and then measuring the direct transmitted beam at $\theta = 0$. A monitor before the sample gave an estimation of the incident flux. A Gaussian function was assumed to described the instrument resolution function which has been proved reasonable in the past [46]. The GRASP (Graphical Reduction and Analysis SANS program) Matlab code package was used for applying a rectangular mask to the direct beam position and then integrate the data radially getting the final curves as intensity (cm^{-1}) versus Q (Å).

Then, using the scattering cross section and the GNOM program [47] the pair distribution function for each temperature was calculated. GNOM approximates $P(r)$ using the regularization method which is based on minimizing the function:

$$T_G(P) = |I(q) - \int_0^{d_{max}} P(r)\frac{sin(Qr)}{Qr}dr|^2 + \mathcal{X}\Upsilon(P) \tag{2.24}$$

where $\Upsilon(P)$ is an stabilizer containing a priori information about the solution. It is normally assume that $\Upsilon(P) = \int \frac{dP(R)}{dr}dr$ which requires $P(R)$ to be a smooth function. $\mathcal{X}$ is the regularization parameter. The larger $\mathcal{X}$ is more attention is paid to the smoothness of the final $P(r)$ and less to fitting the experimental data. The value of $\mathcal{X}$ is crucial and there is not an accepted universal error-proof method to choose the value.

GNOM search for the optima $\mathcal{X}$ using a perceptual criteria. With this criteria the program explore different $\mathcal{X}$ values and estimates the plausibility of the solution by its smoothness, stability with respect to $\mathcal{X}$ variations and the absence of systematic variations.

The perceptual criteria is reported extensively in the book by Tikhonov and Arsenin [48] and well summarized in [49]. The GNOM estimations of $\mathcal{X}$ vary for different SANS curves of a sample at different temperatures. The estimations fall in the range $0.58 < \mathcal{X} < 1.2$. For consistency $\mathcal{X}$ was fixed to 1 for every SANS curve which gives reasonable results for all the data studied.

Regarding the UV-Vis spectroscopy data recorded, since intensity of the UV-Vis radiation through the sample was monitored at each temperature ($I_S(T)$) the absorbance of the sample can be calculated in function of temperature as:

$$A(T) = log_{10}(\frac{L_S - D_S}{I_S(T) - D_S}) \tag{2.25}$$

where L_S is the so called light measurement or the intensity recorded when the sample position is empty (transmittance of 100%), D_S is the dark measurement or intensity recorded when the beam is blocked (transmittance of 0%).

2.5.3 Small angle X-ray scattering

For the SAXS case the reduction of the data is almost identical. The important difference is that the final data was normalized to absolute units using the diffraction of a glassy carbon standard as explained in [50].

The calculation of the $P(r)$ for SAXS data was identical to the one of SANS.

The configuration of the SAXS instrument had some advantages over the D22, for example the increased flux will allowed to perform the experiment with a fraction of the sample used for neutrons. However, there were not in-situ UV-Vis absortion measurements of the samples studied by SAXS.

Chapter 3

Literature review

This chapter presents literature results which are relevant to this thesis as well as some of the models that will be use to analyze and interpret the data.

3.1 Previous diffraction studies in fiber DNA

As described in the comprehensive book by Watson [51] X-ray diffraction studies on DNA fibers by Wilkins and Gosling showed that the DNA molecule can assume a highly periodic structure given that the fibers are kept moist. Frankling and Gosling were the first to develop optimal techniques to control the relative humidity of fibers [52] and they could obtain mode defined diffraction patterns which they designated A and B. They already detected that the pattern of a given fiber changes reversible when the water content of the sample is modified and that mixtures of A and B patterns exits which correspond to samples with a conformational mixture.

Watson and Crick developed a two stranded model [53] which was able to explain the B diffraction pattern and later a more detailed analysis by Wilkins et al. [54] showed that both A and B patterns could be accounted for by the models of watson and Crick.

Since then, many groups greatly contributed to the overall knowledge of the characteristics of crystalline and semicrystaline fiber DNA and their A and B diffraction pattern as well as the A-to-B transition, a non complete list includes: Fuller et al. [55, 41]; Franklin and Gosling [56]; Langridge et al. [38]; Forsyth et al. [57]; Lindsay et al. [58] among others.

Several comprehensive overviews in the topic can be found (e.g. [59]), the rest of the section will be focused in specific results that are specially relevant since they come from very similar experiments performed on almost identical samples with respect to the present work.

3.1.1 Conformation of humidified fibers. A and B form DNA as seen by neutron scattering

As described in sec. 2.3.4 and above the diffraction pattern of A-form and B-form DNA differ dramatically and changes in the water content or hydration of the fibers can trigger a conformational B-to-A transition which is reversible. The conformation of the DNA under study must be taken into account for a correct interpretation of the scattering features.

Valle-Orero et al. [60] studied recently highly oriented wet spun fibers humidified at several relative humidities using neutron scattering. They identified the main features which indicate a given sample is in B form, A form or a mixture of both conformations.

Fig. 3.1 reproduced from [60] summarizes these results. The figure shows selected parts of the reciprocal space maps in longitudinal (fiber axis within the scattering plane) orientation. The maps represent the diffracted intensity in reciprocal space using the perpendicular coordinate system explained in sec. 2.5.1:(Q_H, Q_K, Q_L).

Fig. 3.1a top shows a singular intense Bragg peak centered at $(1.87, 0, 0)$ Å^{-1} which is characteristic of the B-form. This Bragg peak, associated with a d-spacing ≈ 3.4 Å, is related with the long range order of the position of the base pairs along the molecular axis (base pair stacking). This feature will be the focus of the temperature dependence study by

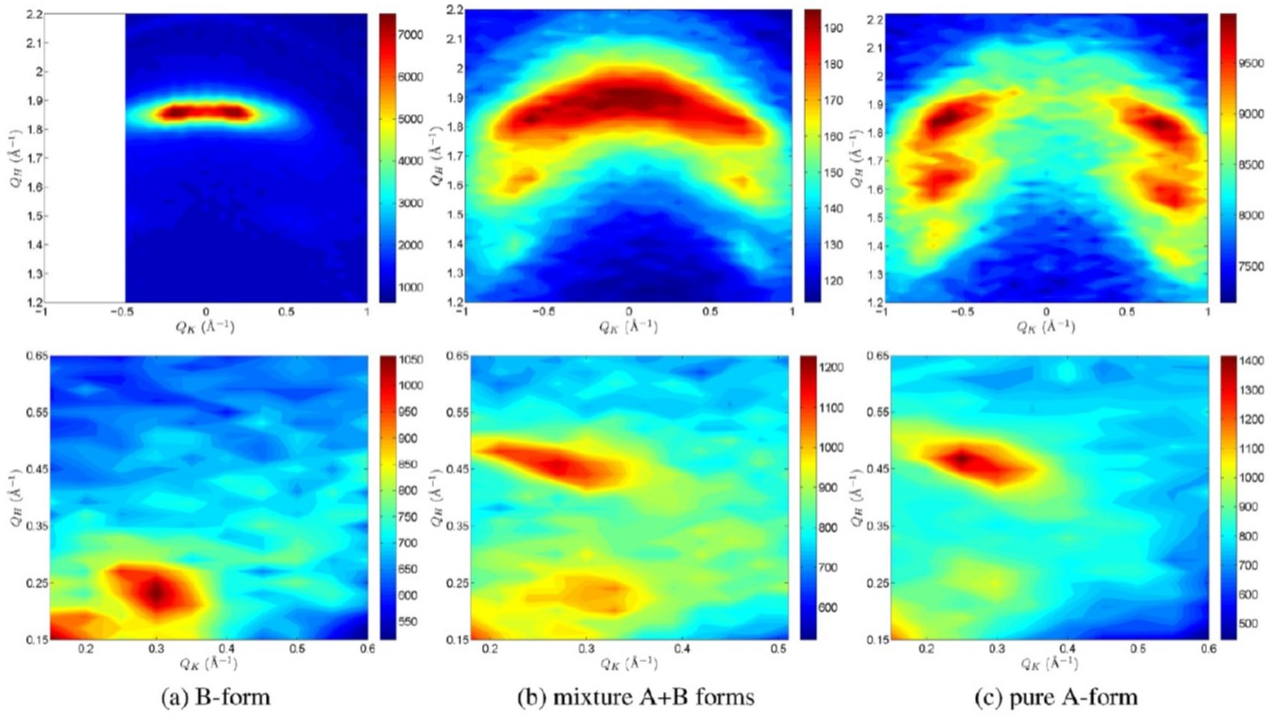


Figure 3.1: Reciprocal space maps in the longitudinal orientation (Q_H-Q_K plane) for a) the B-form, b) a mixture of A+B forms, and c) the A-form. Reproduced from [60].

neutrons in next chapters and using an analogy with X-rays (sec. 2.3.4) it will be referred to as as 10^{th} layer peak.

Fig. 3.1a bottom shows the low-Q part of the longitudinal reciprocal space map for the same sample. The halo of the direct beam is visible in the left bottom corner. The Bragg peak centered at $(Q_H, Q_K) = (0.25, 0.3)$, was identified in [60] as another reflexion related with the B-form.

Fig. 3.1c top shows the high-Q part of the map for the A-form DNA. The three off axis peaks characteristic of the A-form 2.3.4 are clearly visible. Fig. 3.1c bottom shows that for this conformation the peak at $(Q_H, Q_K) = (0.25, 0.3)$ is not present but a Bragg at $(Q_H, Q_K) = (0.45, 0.25)$ appears.

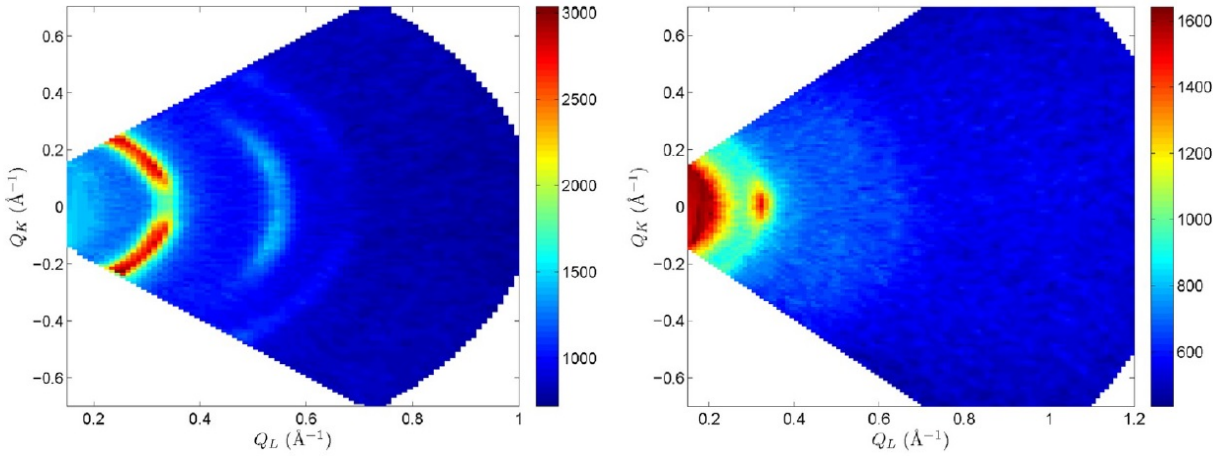
Lastly, in fig. 3.1b top and bottom the mixture of conformations is patent since the features of both B and A form are superimposed.

Note that a pure B-form is desirable for the investigations of the present work since it is biologically more inreresting. It is also easier to analyze than the A-form since it has only one strong peak at large Q rather than 3 off axis peaks.

Valle-Orero et al. also measured the pure B and A form samples in transversal orientation (fiber axis perpendicular to the scattering plane) to obtain recirpocal space maps in the Q_L-Q_K plane. These results are reproduced from [60] in fig. 3.2.

Fig. 3.2a shows two ring-like features: one strong and centered around $Q = 0.5$ Å^{-1}; and another much weaker at around $Q = 0.57$ Å^{-1}. These Bragg peaks are another characteristic feature of the B-form conformation. They prove that there is a semi-crystalline arrangement of the molecules perpendicular to the helical axis as discussed in [23]. These reflexions arise from correlations in the direction perpendicular to the molecular axis. Therefore they are defined by the packing structure of the molecules in the fiber and can be used to gauge the intermolecular distance in a given sample.

Fig. 3.2b shows the pattern related with fibers in pure A-form conformation/ The fig-



(a) (b)

Figure 3.2: Reciprocal space maps in the transversal orientation (Q_L-Q_K plane) for a) the B-form, b) the A-form. Reproduced from [60].

ure does not show ring-like features but a strong localized Bragg peak centered at around $(Q_L, Q_K) = (0.3, 0)$.

3.1.2 Conformation of DNA fibers submerged in ethanol solutions. The B-to-A transition

Zimmerman and Pheiffer proved that the B-to-A transition occurs when the fibers are submerged in ethanol water mixtures with relatively high ethanol content [61]. They used X-ray diffraction to find that *Na*-DNA are in B-form when submerged in ethanol solutions in the range 0-70% ethanol (v/v) and in A-form in the range 75-85%. For ethanol concentrations higher than 85% the diffraction patterns do not resemble the A-form and becomes progressively blurred. This indicates the presence of a less ordered structure.

Zimmerman and Pheiffer also proved [62] that *Li*-DNA samples submerged in ethanol water mixtures remain in the B-form until ethanol concentrations up to 70%. For concentrations higher than 70% a transition to the C-form occurs. The C-form differs from the B-form in that the number of base pairs per helical turn is not as well defined (it is said to

have between 8.8 and 9.7 base pairs per helical turn); the base pairs are also highly tilted.

3.1.3 DNA fibers submerged in PEG solutions. The osmotic pressure method

DNA in PEG solutions have been widely used as a model system of the behavior of DNA in crowded environments. The interest arises because they may reproduce conditions within the cell ([63, 64, 65]). However, to the best of our knowledge, none of the previous studies from other research groups gave information about the spatial correlation along the molecules in this crowded conditions.

The most relevant literature for the present work done on PEG and DNA has been done by Parsegian and coworkers. They have published many articles ([10, 11, 66]) based on the realization that when submerging condensed DNA (either oriented or amorphous aggregates) in PEG solutions the average interaxial distance between DNA molecules is controlled by the osmotic pressure of the solution; which in turn is controlled by the concentration of the dissolved PEG. Given that the PEG is excluded from the DNA fibers, the osmotic pressures acts analogously to a mechanical and permeable piston preventing the fibers from dissolving and allowing the DNA fibers to equilibrate with the solution.

This provides an elegant and simple way of studying the melting transition as a function of the DNA confinement or DNA-DNA interactions.

Note that knowing whether the PEG molecules penetrate the DNA fibers is an important point not only for controlling the intermolecular distance but also for understanding the effect of the PEG in the melting transition of fiber samples. Previous works presenting scattering data of DNA fibers submerged in many different polymer solutions have concluded that whether a specific polymer is able to penetrate the DNA phase depends on its length (proportional to molecular weight) and its concentration(e.g. [10],[11]). They found that PEG with molecular weight ≥ 6000 (PEG6k) at concentrations higher than 7% w/w should be excluded from the DNA fibers [67].

3.2 Confinement of DNA

The question of how the confinement in the aggregated DNA fibers affects the melting transition lies at the heart of the present work. We are interested in how the intermolecular interactions (which depend on the interaxial spacing) change the parameters of the melting transition. Kornyshev and Leikin published a paper which is very relevant to the present work [68]. In this article, they explain a model which allows the force experiencecd by two parallel double helices in columnar aggregates.

They study the interaction between two relatively long, parallel DNA molecules in aqueous solution. In their model each molecule v of length L has a cylindrical, water-impermeable core and a discrete distribution of charges located on a coaxial cylindrical surface around the core of maximum radius R_a. The charges, which are distributed on an helix, are responsible for the interaction. The surface charge density distribution for a given molecule was described in cylindrical coordinates with respect to the molecular axis, $\sigma_v(z, \Phi)$. $\sigma_{0,v}$ is the mean surface charge per unit of area. If the helical strands are continuous, both molecules have similar number and distribution of charges and the adsorbed countercharges are homogeneously smeared at the water-molecule interface the surface charge density for each molecule is [68]:

$$\sigma_v(z, \Phi) = \sigma_{v,0}[2\pi \sum_{i=1}^{2} p_{v,c}\delta(\Phi - \Phi_{v,i} - \frac{2\pi z}{P}) - \theta_v] \tag{3.1}$$

here $i = 1, 2$ labels the strands of each helix, $0 \leq p_c \leq 1$ is the fraction of charge on each strand, $0 \geq \theta_v \leq 1$ is the fraction of DNA charge neutralization by adsorbed cations(a realistic value is $\theta = 0.8$ for any DNA molecule [69]). P is the helix pitch, $\Phi_{v,i}$ is the polar angle of the strand i on the molecule v at $z = 0$.

The electrostatic short-range force field of the system was described with a generalized mean field formalism [70]. Within this formalism the free energy (E_f) of the two-molecular system is expressed through a potential V_a as:

$$E_f = \frac{1}{2}\int \rho_a(\vec{r})V_a(\vec{r})d^3\vec{r} \tag{3.2}$$

where d^3r indicates an integration over the system volume, ρ_a is the volume density of fixed surface changes which is related with the surface charge density by:

$$\rho(\vec{r}) = \sigma(z,\Phi)\delta(I_{ad} - R_a) \tag{3.3}$$

where I_{ad} is the interaxial distance between molecules. V_a satisfies the Poisson equation inside the inner cores: $\Delta V_a(\vec{r}) = 0$. Outside the cores the potential follows the linear Poisson-Boltzmann equation:

$$\Delta V_a(\vec{r}) - \kappa^2 V_a(\vec{r}) = 4\pi\alpha\rho(\vec{r}) \tag{3.4}$$

where κ is the inverse of the Debye length and α is the inverse of the dielectric constant of water.

The average radial force per unit length between the molecules can be calculated using:

$$f = -<\frac{\partial(E_{int}/L)}{\partial I_{ad}}> \tag{3.5}$$

The <> symbols denote averaging over the surface charge density at a fixed I_{ad}. $E_{int}(I_{ad}) = E_f(I_{ad}) - E_f(\infty)$ is the interaction energy for a given surface charge distribution.

The force can experienced by each molecule can be divided in three components:

$$f = f_{cyl} + f_{self} + f_{cross} \tag{3.6}$$

f_{cyl} corresponds to the interaction of two homogeneously charged cylinders, f_{self} is related with the correlated discrete surface charge distribution of each molecule and f_{cross} is due to

a non random alignment of discrete charges on the opposite molecules.

Expressions for the three components can be approximated if the two molecules are assumed to be identical although they still may rotate around its own axis independently. The whole mathematical derivation is long and can be found in [68] (part C, double helices). The final results are:

$$\frac{f_{cyl}}{16\pi^2\alpha\sigma_0} = \frac{1}{2\kappa}\frac{K_1(\kappa I_{ad})}{K_1(\kappa R_a)K_1(\kappa R_a)} \tag{3.7}$$

with $K_n(x)$ being the modified bessel function of the second kind and order nth.

The selfcorrelation term:

$$\frac{f_{self}}{16\pi^2\alpha\sigma_0^2} = \sum_{n=1}^{\infty}\cos^2(\frac{n\Phi_s}{2})\frac{\Omega^{'}_{n,n}(\kappa_n I_{ad}, \kappa_n R_a)}{\kappa_n[K^{'}_n(\kappa_n R_a)]^2} \tag{3.8}$$

where $\kappa_n = \sqrt{(}\kappa^2 + (\frac{2\pi}{P})^2 n^2)$, $K^{'}_n(x) = dK_n(x)/dx$ and Φ_s is the azimuthal half-width of the minor groove (which is $\approx 0.4\pi$ for B-DNA [69]) and $\Omega^{'}_{n,n}$ is so that:

$$\Omega^{'}_{n,m}(x,y) = \sum_{j=-\infty}^{\infty}\left([K^{'}_{n-j}(x)K_{j-m}(x) + K_{n-j}(x)K^{'}_{j-m}(x)]\frac{J^{'}_j(y)}{K^{'}_j(y)}\right) \tag{3.9}$$

where J_n is the bessel function of the first kind order n^{th} and $J^{'}_n(x) = dJ_n(x)/dx$.

$$\frac{f_{cross}}{16\pi^2\alpha\sigma_0^2} = \sum_{n=1}^{\infty}(-1)^n\cos^2(\frac{n\Phi_n}{2})\cos[n(\Phi_1 - \Phi_2)]\frac{K_1(\kappa_n I_{ad})}{\kappa_n[K^{'}_n(\kappa_n R_a)]^2} \tag{3.10}$$

where $(\Phi_1 - \Phi_2)$ is the relative rotation of the two identical molecules. For the calculations of the present work $(\Phi_1 - \Phi_2)$ was taking as the optimal relative angle which minimizes the cross correlation energy which can be calculated as:

$$(\Phi_1 - \Phi_2) = \arccos\Big(\frac{\kappa_2^2 \cos^2(\frac{\Phi}{2})[K_2^{'}(\kappa_2 R_a)]^2 K_0(\kappa_1 I_{ad})}{a\kappa_1^2 \cos^2(\phi_s)[K_1^{'}(\kappa_1) R_a]^2 K_0(\kappa_2 I_{ad})}\Big) \tag{3.11}$$

this approximation holds for inter axial distances shorter than a critical separation of R^*, at larger distances the optimal angle is zero. R^* can be approximated as:

$$R^* = \frac{1}{\kappa_2 - \kappa_1} ln\big[\frac{4\cos^2(\Phi_s)[K_1^{'}(\kappa_1 R_a)]^2 \kappa^{5/2}}{\cos^2(\Phi_s/2)[K_2^{'}(\kappa_2 R_a)]^2 \kappa_2^{5/2}}\big] \tag{3.12}$$

Forces calculated with this model were proven to be consistent with experimental measurements form Parsegian and coworkers [67].

Note that under certain conditions this model predicts an attraction force between DNA molecules. DNA molecules are generally negatively charge in solution so this may look unreasonable. In a different publication Kornyshev and Leikin proved that an attractive resultant force can take place caused by an attraction of the phosphates of strands of one DNA with cations adsorbed in the grooves of another DNA when DNA is confined to dense aggregates [71].

3.3 Melting transition

The melting transition is usually characterized by the melting temperature, T_m. T_m is defined as the temperature at which half of the base pairs in the molecule are open. It is commonly used for addressing the stability of DNA and other nucleic acids [72, 73]. In long chain DNA the melting temperature depends mainly on the environmental conditions of the DNA like hydration level, ionic strength of the medium, the presence of other solutes in the surrondings and also the DNA-DNA interactions.

Another feature that helps characterizing the transition is the cooperativity. The cooperativity of the transition is the tendency of the base pairs to open in contiguous segments. The cooperativity is usually quantified by the so called cooperativity factor which is the mean

number of base pairs that melt as a single thermodynamic entity [74]. In a highly cooperative transition, usually shown by short DNA molecules and specially synthetic polynucleotide duplexes (poly(dAdT) or poly(dGdC)), most of the base pairs of the molecule open as a whole at a single temperature. In a less cooperative transition, ascribed traditionally to long natural DNA, many domains in the molecules melt independently from each other at different temperatures. The cooperativity depends on the energy and entropy changes associated with the creation of the open regions between helical domains.

The cooperativity can be accessed experimentally since the width of the melting transition (which can be measured, for example, by calorimetry as explained in sec. 2.2.5) is inversely proportional to its cooperativity. Thus narrower transitions are more cooperative [75].

3.3.1 Effect of the confinement of DNA in the melting transition

A prediction of how the intermolecular forces in dense DNA assemblies affect the melting transition was published by Cherstvy and Kornyshev [69]. They combined the theory of electrostatic interactions presented in sec. 3.2 (also see [68]) and the two-state one-dimensional Ising model of DNA melting [76].

The model was used to account for the difference in the free energy of base pairs in closed (F_c) and open (F_o) state. During the melting of a single molecule: $F_o - F_c = \Delta U - T\Delta S$ where T is temperature, ΔU is the heat of melting of a base pair and ΔS is the entropy difference between closed and open states. $\Delta S > 0$ because open base pairs have more degrees of freedom. At the melting temperature of this single molecule, T_{m0}, the number of open (N_o) and closed (N_c) base pairs is equal so:

$$s = exp[(F_o - F_c)/(k_B T_{m0})] = 1 \tag{3.13}$$

where k_B is the Boltzmann constant.

This model also includes a cooperativity factor $\sigma_{coop} = exp[-F_s/(k_B T)]$ where F_s is the

energy required to create a open region between two intact closed regions. Smaller σ_{cop} values (larger F_s) corresponds to a more cooperative melting.

Cherstvy and Kornyshev argued that if the force between DNA molecules is attractive this attraction stabilizes the close base pairs, thus it lowers the free energy in the closed state. Therefore the melting temperature must increasse for eq. 3.13 to still hold [69].

Following previous studies in the topic [77, 78] they modeled the interaction energy between parallel soft non-homologous DNA molecules (totally independent sequences) as:

$$E(L) = a_0 L - L a_1 [1 - l_t/(2l_c)] - a_1 l_t^2 [1 - e^{l/l_t}]/(2lc) \tag{3.14}$$

where L is the length of the molecules, l_c is the helical coherent length (which has a value between 300-700 Åfor DNA [79, 80]), $l_t = \sqrt{C_t/2a_1}$ and C_t is the torsional elastic modulus of DNA (usually taken as 3×10^{-18} J Å [81]) . $a_0(I_{ad})$ and $a_1(I_{ad})$ are the electrostatic interaction harmonic coefficients whose expressions can be found in the appendix of [69]. As in previous sections, I_{ad} is the interaxial distance between molecules.

Using eq. 3.14 Cherstvy and Kornyshev rewrote the free energy of the Ising model as the sum of the interaction energy of the closed fragments, the energy of the closed/open boundaries and the energy of all the DNA base pairs which allowed them to obtain an expression for the free energy as a function of N_o, N_c, T, I_{ad} and n_b which is the number of closed/open boundaries. Minimizing F with respect n and N_h. They derived two coupled equations which must be fulfilled at the melting temperature in a columnar DNA aggregate:

$$\begin{aligned} (\frac{N}{2n_b} - 1)^2 &= \frac{1}{\sigma_{coop}} \\ \frac{1 - 2_b/N}{2(1 - n_b)/N} &= exp(-\frac{1}{2}\frac{6a(a_0 - a_1)}{k_B T}) \end{aligned} \tag{3.15}$$

where a is the distance between base pairs (3.4 Åfor B form), n_b is the number of closed/open boundaries and $N = N_c + N_o$ is the total number of base pairs.

If all the other parameters are constant given two interaxial distance: I_{ad1} with melting temperature T_{m1} and I_{ad2} with melting temperature T_{m2} this model predicts that :

$$T_{m2} = T_{m1} \frac{a_0(I_{ad2}) - a_1(I_{ad2})}{a_0(I_{ad1}) - a_1(I_{ad1})} \tag{3.16}$$

3.3.2 Effect of PEG in the melting transition

To the best of our knowledge there are not previous studies on the melting of DNA fibers submerged in PEG solutions. However, the melting of DNA dissolved in many different PEG solutions was reported by different groups (e.g. [82, 83]) and the results are relevant to the present work.

When PEG is dissolved in a DNA-water solution it decreases the water activity of the solution which in turn decreases the melting temperature of DNA because it changes the water structure around the DNA molecules [82]. This is believed to be the explanation for low-molecular weight PEGs and other polymers decreasing the melting temperature of DNA dissolved in solution. However, high molecular weight polymers increase the melting temperature of DNA in solution. The origin of such increase was said to be the excluded volume effect. This assumes that the volume of the solution unavailable to the DNA due to the PEG presence stabilizes the molecular conformation which occupies the smallest space, that is, the helix molecule [82, 83]. The decrease of the water activity still happens for the long chain polymers but its effect on the melting has been proven to be very small in comparison to the effect of the excluded volume [82].

A strong indication that the excluded volume effect is taking place is that the variation of T_m is changing dramatically with the molecular weight of the PEG [82, 73], as can be seen in fig. 3.3.

Even though it is not patent, these results for DNA in solution are relevant for the case of DNA-fibers submerged in PEG solutions. If the PEG penetrates the fibers then the

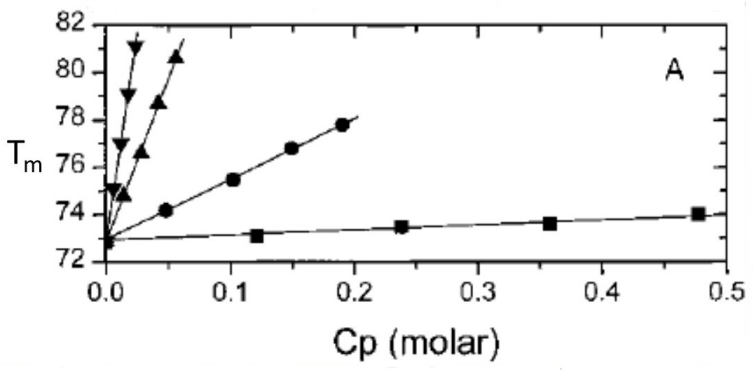
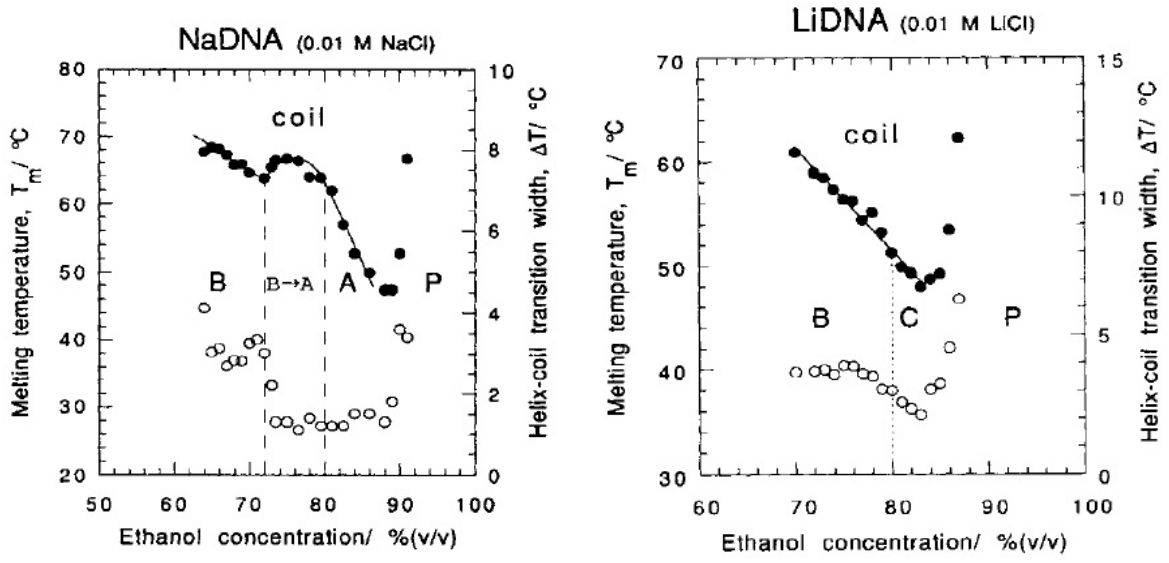


Figure 3.3: Variation of T_m of *E.coli* DNA in solution with the molar concentration of the PEG 400 (squares), 1000 (circles), 3400 (triangles), 8000 (inverse triangles). Reproduced from [82].

sample is not much more different than a very concentration DNA solution and the excluded volume of the PEG will affect the melting transition of the fibers. On the contrary, if the PEG remains separated from the DNA, in the outside of the fibers, just a small fraction of the DNA molecules in the sample will be in contact with PEG molecules and the excluded volume effect will be negligible.

3.3.3 Effect of ethanol in the melting transition

It is known that increasing concentrations of ethanol decrease the melting temperature of DNA in solution [84]. The main factors explaining this behavior are the increased electrostatic repulsions among phosphate groups of the helix due to the lowering of the dielectric constants of the ethanol solution as the ethanol concentration increases, the hydrophobic effect and the decreased water activity [84]. In the case of the fibers submerged in ethanol solutions there is not a phase separation as happens with PEG of a high enough molecular weight. Ethanol is a very small molecule in comparison with PEG (the molecular weigh of ethanol is 46.07 g/mol) and will penetrate the fibers as easily as the water molecules. There-



(a) (b)

Figure 3.4: Melting temperature (filled circle) and width of the transition (open circle) for fiber submerged in ethanol solutions. a)Na-DNA and b)Li-DNA. Reproduced from [85].

fore the same factors which affect DNA in solution will for sure be present in the submerged fiber samples.

Rupprecht et al. studied the melting transition of very similar DNA wet-spun fibers samples submerged in ethanol/water mixtures ([85],[86]). They used a mechanochemical method which can relate changes in the length of the submerged fibers strained by a weight with confomational and thermodynamical transitions of the DNA [87]. Fig. 3.4 summarizes their results for $Na-$ and Li-DNA fibers.

For Na-DNA (fig. 3.4a) T_m decreases from around 60% to 88% ethanol concentration ([EtOH]) but there is a local maximum at around 75%. The width of the transition decreases slightly from 60% to 70% and drops steeply at around 75% [EtOH]. These seemingly coupled changes in the behavior of T_m and the width were related with the B-to-A form transition of the secondary structure of the DNA (section 2.3.4). Rupprecht et al. linked the effect of the B-to-A transition on T_m and width to the stronger inter-helical interactions that the A-form exhibits with respect to the B-form in aggregated DNA [88].

For [EtOH] higher than 90% the changes in T_m and width were related with a transition to P-form DNA [89], a very stable and dehydrated DNA form which is not of interest for the present work.

For *Li*-DNA fibers (fig. 3.4b) the T_m decreases monotonically until around 83%. The width stays constant from 70% to 75% and then decreases steadily. The raise in T_m and width after $\approx$ 82% is again related with the transition to the P-form. In these samples a B-to-C transition was detected at around 80% ethanol.

These transitions were also detected by X-ray scattering in [61] (B-to-A) and in [62] (B-to-C) as explained in sec. 3.1.2.

Rupprecth et al. measured samples submerged in ethanol solutions in a rather extensive range (0.3-3 M) of salt concentrations ($NaCl$ for Na-DNA fibers and $LiCl$ for Li-DNA fibers). They found no variance of the melting transition with salt concentration in this range [87].

3.4 Models of DNA and the melting transition

3.4.1 Peyrard-Bishop-Dauxois model

The one dimensional Peyrard-Bishop-Dauxois (PBD) model [90], [91] was used to analize the part of the neutron data. The model is able to predict melting curves of complex DNA molecules and solve the dynamical properties associated to the formation and stability of the denaturation bubbles [92].

A simple version of the model describes a DNA molecule as N base pairs placed in a 1D disordered chain with a mean separation of a, fig. 3.5. The state of each base pair is described by a single real variable, y_n, which is the stretching of the base pair perpendicular to the molecular axis with respect to its equilibrium position. y_n ranges from 0 to $+\infty$. The Hamiltonian of such a system is:

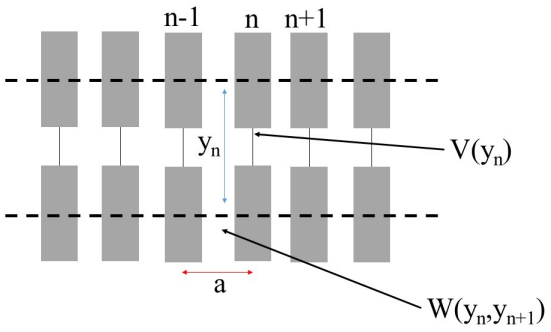


Figure 3.5: Graphical representation of the PBD model. y_n is the stretching of the n^{th} base pair, a is the mean distance between consecutive base pairs, x_n is the local structure deviation for the n^{th} base pair, W is a potential describing inter-pair interaction and V is a potential accounting for the interactions between the two bases within the pair n^{th}.

$$H_y = \sum_{j=1}^{N-1} W(y_j, y_{j+1}) + \sum_{j=1}^{N} V_j(y_j) \tag{3.17}$$

where y_j represents the stretching of the j^{th} base pair caused by the transverse displacement of the bases. $W(y_j, y_{j+1})$ is a potential describing the stacking interaction between adjacent bases [93]:

$$W(y_j, y_{j+1}) = \frac{1}{2}k[1 + \rho e^{-b(y_j + y_{j+1})}](y_j - y_{j+1})^2 \tag{3.18}$$

where k is a coupling constant, ρ is the relative strength of the nonlinear base stacking and b is the inverse range of the nonlinear base stacking. V_j describes the intra-pair potential in the j^{th} base pair and takes into account the hydrogen-bonding, electrostatic interactions between phosphate groups and solvent effects. It is usually taken as a Morse potential:

$$V_j(y_j) = D_j[1 - e^{-\alpha_j y_j}] \tag{3.19}$$

with D_j and α_j being respectively the depth and inverse range of the potential. Four

different parameters were necessary for describing the two kinds of base pairs:D_{AT}, α_{AT} and D_{GC}, α_{GC}.

The model can be used to calculated the size distribution of the closed segments of DNA which would be instrumental for the analysis of the diffraction data presented in this work. A detailed description of this calculation has been reported elsewhere [7]. Here only main steps of the process are presented.

A quantitative definition of a closed base pair is needed to calculate the size of the closed regions at a given temperature (T) . Base pair j is considered closed if $y_j < y_c$ where y_c corresponds to a stretching well on the plateau of the Morse potential.

The statistical weight of a given configuration of the molecule thermalised at temperature T is[7]:

$$z_T = \int_{y_{l1}}^{y_{f1}} dy_1 \cdots \int_{y_{lN}}^{y_{fN}} dy_N \cdot exp\big(- H_y(y1, \cdots, y_N)/k_B T\big) \tag{3.20}$$

where k_B is the Boltzmann constant and y_l and y_f define the configuration. If $y_l = -\infty$ and $y_f = \infty$ for all j, the molecules explore the whole configurational space and $z_T = Z$ where Z is the partition function. $y_{lj} = 0$ and $y_{fj} = y_c$ define a configuration in which the base pair j is closed. If $y_{lj} = y_c$ and $y_{fj} = \infty$ define a configuration in which base pair j is open. The integrals associated with all configurations can be easily calculated as a series of mono-dimensional integrlas because the model is one-dimensional and restricted to nearest neighbor coupling. [7]

The probability of m adjacent base pairs being closed starting with the site i is then:

$$\mathcal{P}(m, i) = z_T(y_1, \cdots, y_{i-1}, y_i < y_c, y_{i+1} < y_c, \cdots, y_{i+m-1} < y_c, y_{i+m}, \cdots)/Z \tag{3.21}$$

Eq. 3.21 calculates the statistical weight of configurations where restrictions on the integration range were imposed for all sites belonging to the closed region and no restrictions

were imposed elsewhere.

The probability of a closed region of size m (with open ends) exist starting at size i is:

$$P(m,i) = \mathcal{P}(m,i) - \mathcal{P}(m+1,i-1) - \mathcal{P}(m+1,i) + \mathcal{P}(m+2,i-1) \tag{3.22}$$

Therefore the probability of having a closed cluster of m base pairs in a DNA segment of length N base pairs is:

$$P(m) = \frac{1}{N}\sum_{i=1}^{N} P(m,i) \tag{3.23}$$

Regarding the selection of this model over others: Other models for DNA denaturation, like the Poland-Scheraga model, are Ising based models. Thus they treat a base pair as a two-system state (closed or open) [76]. Even though such models are appealing because of their simplicity they need a great number of phenomenological parameters which may not always be easy to determine. The PBD model contains many simplifying assumptions but also few free parameters that can be calibrated from experimental data. The model not only allows fast calculations but also includes some structural information in the stretching variable [7].

Calculation of the $S(Q)$ of an oriented DNA molecule

The PBD model can be used to calculated the expected diffraction pattern ($S(Q)$) of a given sequence of DNA which will be instrumental for the present work. The procedure to calculated $S(Q)$ has been extensivily reported elsewhere [7, 94]. A summary will be presented here.

Wildes et al. developed an expression for the structure factor of a structurally disordered linear chain of m sites with average length between the sites of a [7, 94]:

$$S_m(\vec{Q}) = m + 2\sum_{n-1}^{m-1}(m-n)\cos(Q_H n a)e^{-n\Delta} \tag{3.24}$$

where the exponential therm is analogous to a Debye-Waller factor which accounts for structural and thermal disorder. Given that $Q_K = 0$, which is the interesting case for the present study, $\Delta = Q_H^2(<\Sigma^2> + <\Lambda^2>)/2$.

$<\Sigma^2>$ is proportional to the longitudinal thermal motions of the DNA molecule and was estimated using $<\Sigma^2> = k_B T/\mu c_0^2 a$ as done in [7] , where k_B is the Boltzmann constant, $\mu = 618\ a.m.u$ is the DNA mass per base pair and $c_0 = 2830\ ms^{-1}$ is the sound velocity in DNA measured in [95].

$<\Lambda^2>$ is the standard deviation of the distance between consecutive base pairs and accounts for the structural disorder along the molecular axis.

It is important to note that near the 10^{th} layer line base peak (where this study is focused) and for m sufficiently large $S_m(\vec{Q})$ can be approximated as $S_m^*(\vec{Q})$:

$$S_m^*(\vec{Q}) = mS'(\vec{Q}) = m\frac{\sinh\Delta}{\cosh\Delta - \cos(Q_H a)} \tag{3.25}$$

For a natural DNA sample with M base pairs and consisting of open and closed domains the observed intensity in a scattering experiment will be proportional to [7, 94]:

$$S(\vec{Q}) = \sum_{m=1}^{M} P(m)S_m(\vec{Q}) \tag{3.26}$$

where $P(m)$ is given by eq. 3.23. The calculations presented in this work were performed with DNA segments of $M = 10000$. Calculating $P(m)$ directly for segments of this size would be unfeasible. Thus $P(m)$ was computed directly with eq. 3.23 for sizes up to $M_0 = 150$. For $m > M_0$, $P(m)$ can be calculated as $P(m) = P_0\zeta^m$ [7] where P_0 is the probability of two consecutive base pairs being in the close state and ζ results from fitting $P(m)$ by straight lines for $40 < m < M_0$ as done in [92]. $S(\vec{Q})$ was finally approximated as:

$$S(\vec{Q}) = \sum_{m=1}^{M_0} P(m) S_m(\vec{Q}) + \sum_{m=M0+1}^{M} P_0 \zeta^m S_m(\vec{Q}) + S'(\vec{Q}) P_0 \zeta^{M_0+1} [\frac{M_0}{1-\zeta} + \frac{1}{(1-\zeta)^2}] \quad (3.27)$$

where $S_m(\vec{Q})$ was calculated with eq. 3.24 for $m < 1000$ and with 3.25 for $m > 1000$.

3.4.2 Kratky-Porod model

The Kratky-Porod model in polymer physics is used to describe the behavior of a semi-flexible polymer. It was developed originally by Kratky and Porod [96]. The polymer is described as a chain of N segments of length a and orientation vector t_i. Simple versions of the model do not take into account excluded volume or torsional stress. In such a case the energy of a given chain configuration, E_{KP}, is:

$$E_{KP} = -\frac{B}{b} \sum_{i=2}^{N} t_i \cdot t_{i-1} = -\frac{B}{b} \sum_{i=2}^{N} \cos(\theta_i) \quad (3.28)$$

where θ_i is the angle between t_{i-1} and t_i and B is the bending modulus of the molecule. The correlation between angles along the chain decays exponentially with distance:

$$< t_i \cdot t_j >= e^{-a|i-j|/l_p} \quad (3.29)$$

where $l_p = B/k_B T$ with K_B being the Boltzmann constant and T the temperature. l_p is a decay length for the correlation which is known as the persistence length.

For a chain with some degree of stiffness, segments which are adjacent to each other tend to point to the same direction. This is, the local contour tends to persist in a given direction and the segment directions are strongly correlated. This is why l_p is usually used as a measured of flexibility. Polymers with longer l_p are considered stiffer.

Kratky and Porod derived the expresion for the mean-squared end-to-end distance also known as radius of gyration:

$$R_g^2 =< R^2 >= 2l_p^2[L/l_p - 1 + exp(-L/l_p)] \tag{3.30}$$

where $L = Na$ is the countour length of the molecule. R_g accounts for the effect of the flexibility of the molecules in the overall conformation of the ensemble.

Chapter 4

Study of submerged DNA fibers

In this chapter the results of investigating fibers submerged in PEG and ethanol solutions will be discussed. The chapter is divided in the structural characterization performed at room temperature using X-ray and neutron scattering and the study of the melting transition of the samples with calorimetry and neutron scattering.

4.1 Structural characterization with X-rays

X-ray patterns collected in the Genix lab source (section 2.4.2) were used to check the existence of cristallinity and the conformation of DNA fibers submerged in ethanol/^2H_2O mixtures. The ethanol solutions are described in sec. 2.1.1 and they will be identified by the amount of ethanol in volume per cent. Fibers submerged in PEG solutions were also investigated with this X-ray lab source but the patterns were featureless and no useful data were obtained.

As references Na- and Li-DNA fibers equilibrated with a humid atmospheres at 75% relative humidity (RH) were measured. The diffraction patterns are shown in fig. 4.1a and 4.1b where the momentum coordinate system is defined by Q_H vs. Q_K where Q_H is parallel to the molecular axis and Q_K is perpendicular to it (as described in sec. 2.5.1).

The Na-DNA fibers equilibrated at 75% RH (fig. 4.1a) show a pattern characteristic of A-form DNA, as reported previously [97, 23] and shown in sec. 2.3.4. The three off-axis Bragg peaks are clearly visible and the many Bragg peaks around the central part of the image indicates a highly crystalline structure.

The Li-fibers equilibrated at 75% RH (fig. 4.1b) show a pattern characteristic of the crystalline B-form DNA. This is consistent with previous studies [98, 23]. The most remarkable features are the 10^{th} Bragg peak given by the strong correlations between consecutive base pairs and the characteristic X-shaped pattern in the central part which is a footprint of the helical structure of DNA (sec. 2.3.4).

Fig. 4.1c and fig. 4.1d show the diffraction pattern of Na-DNA fibers submerged in ethanol solutions at 60% and 70% respectively. The images at ethanol concentrations below 60% are qualitatively similar to the fig. 4.1c and they are not presented. These data suggest that the Na-fibers submerged in ethanol solutions are in the B-form for concentrations below 60% and in the A-form for ethanol concentrations above 70% with a transition at some point between 60-70% ethanol concentration. This is consistent with previous X-ray studies as presented in sec. 3.1.2 and in [61].

Zimmerman and Pheiffer found the B-A transition between 70-75% ethanol concentration [61] but such a relatively small different can be due to a different $NaCl$ content in the solution used to submerge the fibers. Zimmerman and Pheiffer worked with $[NaCl]$ as low as 0.02 M in comparison to the 0.1 M used in the present work. A higher salt content favors the A-form [58] and so it may decrease the ethanol concentration required for the transition to occur.

On the contrary fig. 4.1e and fig. 4.1f show that Li-DNA fibers exhibit a B-form conformation in the concentration range studied (20-70%) which is again consistent with previous works (sec. 2.3.4 and [62]). Data above 70% where the B-to-C transition is expected to occur were not collected in this work.

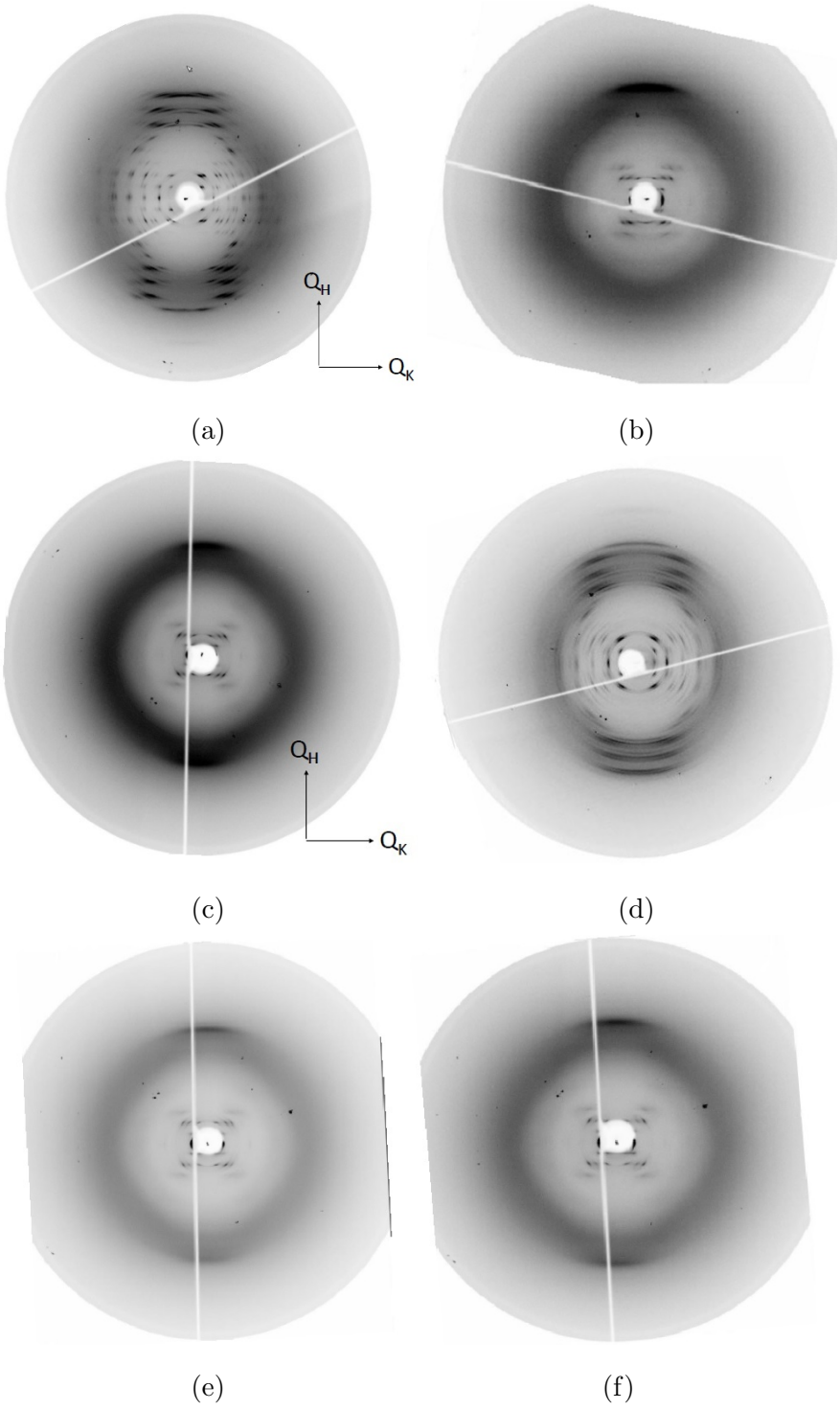


Figure 4.1: X-ray diffraction photographs of DNA fibers. Fibers equilibrated with an atmosphere of 75% RH a) Na-DNA and b) Li-DNA. Fibers submerged in ethanol solutions: c) Na-DNA 60% , d) Na-DNA 70%, e) Li-DNA 20% f) Li-DNA 70%. The momentum coordinate system is defined by Q_H vs. Q_K where Q_H is parallel to the molecular axis and Q_K is perpendicular to it.

4.2 Room temperature study with neutrons

4.2.1 Fibers submerged in PEG solutions

Five samples of Na-DNA submerged in PEG solutions were studied by neutron scattering at room temperature. The solutions contained: PEG 6000 at 17% (w/w) (PEG6K-17%); PEG 6000 at 20% (PEG6K-20%); PEG6K-40%; PEG 8000 at 15% (PEG8K-15%); and PEG 8000 at 20% (PEG8k-20%). As a reference, Na-DNA fibers humidified at 92% RH in a ^2H_2O atmosphere were also measured (Na-humid).

The left panels of fig. 4.2 shows the reciprocal space maps for the Na-humid fibers. Fig. 4.2a is the map of the longitudinal orientation (fiber axis within the scattering plane). Fig. 4.2c is the map of the transversal orientation (fiber axis perpendicular to the scattering plane). The maps represent the diffracted intensity in reciprocal space using the perpendicular coordinate system explained in sec. 2.5.1: (Q_H, Q_K, Q_L).

Na-fibers at this RH are expected to be in the semicristaline B-form of DNA [98, 60]. Indeed, fig. 4.2a shows a singular intense Bragg peak centered at $(1.87, 0, 0)$ Å^{-1} (the 10^{th} layer Bragg peak) which is characteristic of the B-form as discussed previously (sec. 3.1.1).

The transversal reciprocal map shown in fig 4.2c is qualitatively identical to the results presented in 3.1.1 and also suggest that the sample is in the B-form. It shows two ring-like features, one strong and centered around $Q = 0.5$ Å^{-1} and another much weaker at around $Q = 0.57$ Å^{-1}. These reflexions are related to the semicrystalline packing of the DNA molecules in the direction perpendicular to the fiber axis [60, 23].

The fibers submerged in PEG solutions all showed very similar features and only the data of the PEG6k-17% sample are shown in fig. 4.2b and 4.2d as model examples. The submerged samples have substantially more incoherent scattering in comparison to the humid fibers due to the presence of protonated PEG and so the signal-to-noise ratios of the Bragg peaks are smaller. The features may also be broader in the sample rotation angle, suggesting that

some orientation of the fibers has been lost. However, the results are qualitatively identical to the humid fibers. This means that the fibers submerged in PEG solutions are also in the B-form and most of the long range order of the fibers is not lost during the submersion. Similar results have been reported previously by Wildes et al. [12].

Fig. 4.3 shows the low-Q part of the reciprocal space map in longitudinal orientation for the Na-humid and PEG6K-17% fibers. The maps are featureless for these sample except for the halo of the direct beam in the left bottom corner. This result is surprising since previous studies in very similar samples concluded that an additional Bragg peak in this region, centered at $(Q_H, Q_K) = (0.25, 0.3)$, was another reflexion related with the B-form (sec. 3.1.1 and [60]). This featured has not been detected in these or other samples measured in the present work. This suggest that the peak at $(0.25, 0.3)$ was related with some specific crystalline arrangement of the samples studied in [60] but does not seem to be a universal feature of the B-form of DNA.

The data of the transversal reciprocal space maps were used to calculate the interaxial distance between molecules (I_{ad}) for every sample and thus to gauge the effect of the PEG solution on the confinement of the molecules. The data of these maps were summed over the sample rotation angles for equal Q values and plotted as a function of the magnitude of Q (the scattering vector). Fig. 4.4 shows the results.

The peaks for the submerged samples are shifted to lower Q values with respect to those of the Na-humid fibers. This suggest an increase in the lattice parameters which is consistent with the fibers swelling when submerged. B-form DNA fibers are believed to have a hexagonal space group, as shown for B-form Na-DNA at 92% RH by Langridge and Wilson [38] and for samples submerged in PEG solution by Podgornik et al. [11]. The two peaks can be indexed as the 110 and 200 reflexions respectively from the hexagonal lattice. The d-spacing related with a given peak is $d = 2\pi/Q_C$ where Q_C is the center of the peak and the relation between the Miller indices and the lattice parameter a, corresponding to the axis-to-axis distance between two molecules, is

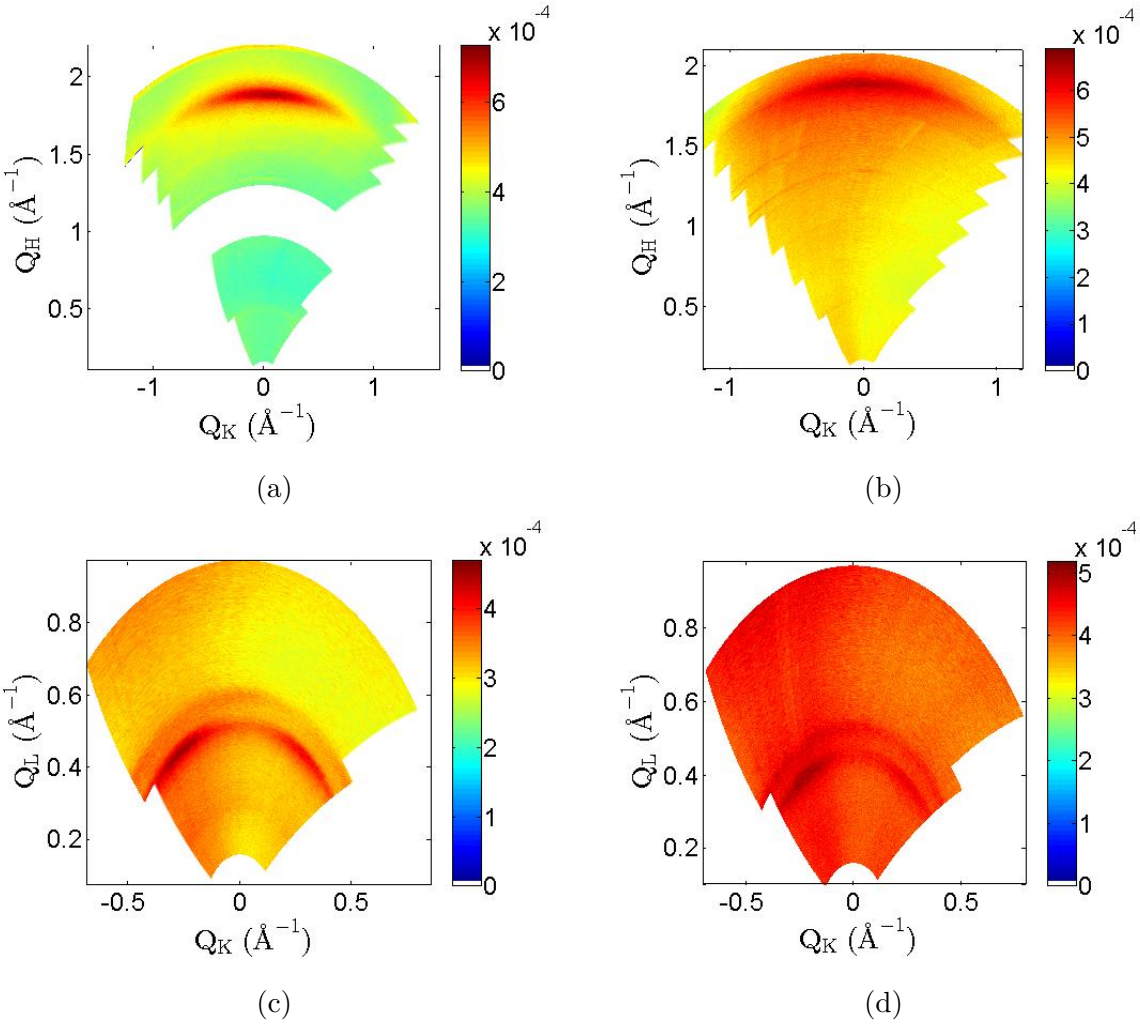
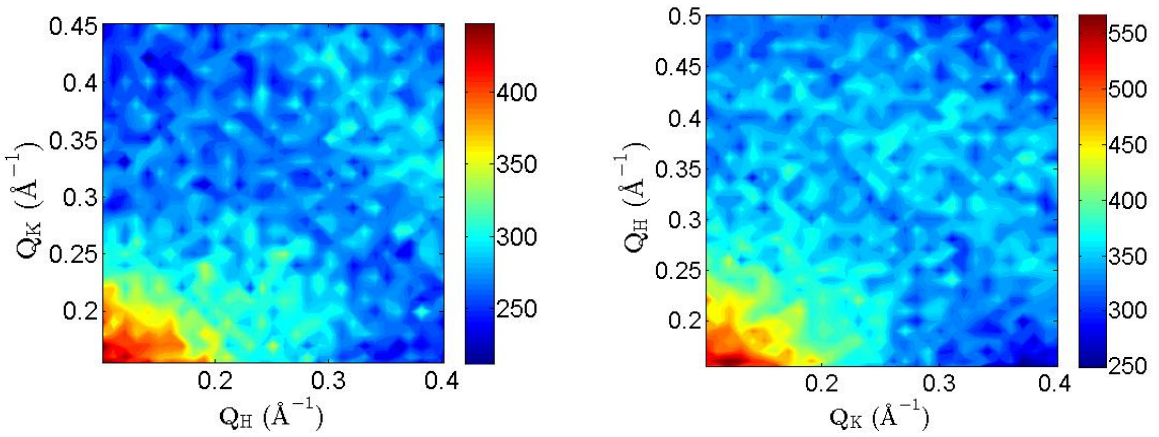
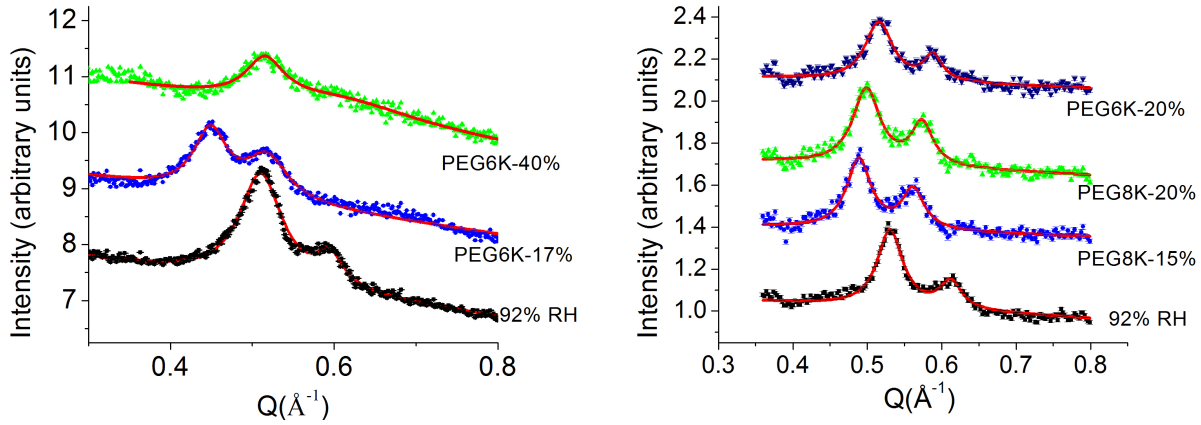


Figure 4.2: Reciprocal space maps of the *Na*-DNA fibers in the longitudinal (top) and transversal (botton) orientations. a) and c) are the maps for fibers humidified to 92% RH. b) and d) are maps for fibers submerged in a PEG6K-17% solution. Data measured in D16.



(a) (b)

Figure 4.3: Low Q-part of the reciprocal space maps of the longitudinal orientation for a) the *Na*-humid fibers and b) the fibers submerged in PEG6k-17%. Data measured in IN3.



(a) (b)

Figure 4.4: Sum of the intensity over the sample rotation angles of the transversal reciprocal space maps for constant scattering vector as a function of Q. The curves have been shifted vertically for clarity. Red lines are the fit to the data. Data on the left were measured on D16, data on the right on WOMBAT.

Sample	I_{ad} (Å)	$\frac{\Delta a}{a_{Hu}}(\%)$
PEG6k-17%	28.24 ± 0.05	14.54 ± 0.91
PEG8k-15%	26.92 ± 0.06	9.18 ± 0.99
PEG8k-20%	26.11 ± 0.04	5.90 ± 0.81
PEG6k-20%	25.48 ± 0.03	3.33 ± 0.70
PEG6K-40%	24.08 ± 0.03	-2.35 ± 0.62
Humid fibers	24.66 ± 0.09	0

Table 4.1: Percentage increase in the intermolecular spacing for Na-DNA fibers submerged in PEG solutions With Respect to the Humidified Fibers.

$$a = d(\frac{4}{3}(h^2 + kh + k^2))^{1/2} \tag{4.1}$$

The centers of the peaks for each sample were determined by fitting two Gaussian functions with a linear background. The results of the fits can be seen in fig. 4.4. The lattice parameter of the Na-humid sample was calculated using the average a from both peaks. The result was $a_{Hu} = 24.66 \pm 0.09$ Å, which is consistent with previous results [38]. The intermolecular distances for the other samples were computed analogously and the percentage increase in intermolecular distance with respect to the humidified fibers was calculated. Table 4.1 shows the values of the percentage increase and the interaxial distance (I_{ad}) for each sample. All the samples swelled under submersion (I_{ad} increased) except for PEG6K-40% in which the osmotic pressure compressed the fibers to I_{ad} shorter than that of the Na-humid fibers.

4.2.2 Fibers submerged in ethanol solutions

Six samples submerged in deuterated ethanol solutions were investigated with neutrons at room temperature: Na-DNA in 60% ethanol-d6 (v/v) with 0.1 M $NaCl$ (Na-60%-0.1); Na-DNA in 66% with 0.1 M $NaCl$ (Na-66%-0.1); Na-DNA in 66% with 0.01 M $NaCl$(Na-66%-0.01); Li-DNA in 20% (Li-20%); Li-DNA in 60%(Li-60%); and Li-DNA in 80% (Li-80%).

All the solutions for lithium samples had a $LiCl$ concentration of 0.1 M and so it has not been included in the code of the sample. As a reference for the lithium samples, Li-DNA fibers humidified at 75% RH (Li-humid) were also measured.

Ethanol mixtures present some advantages over polymer solutions for the studies of DNA fibers using neutron scattering: a) they do not contain long-chain polymers which contribute to the scattering of the sample and could mask the signal of the DNA. b) fully deuterated ethanol is easily accessible while fully deuterated PEG is prohibitively expensive. Minimizing the amount of ^1H in the sample is important for neutron scattering since it contributes to the incoherent scattering and degrades the peak-signal-to-noise ratio.

As shown in sec. 4.1 the Na-fibers submerged in ethanol undergo a B-to-A transition due to the ethanol drastically reducing the water activity of the solution [61].

Fig. 4.5 shows a longitudinal reciprocal space map for Na-66%-0.1. It shows off-axis peaks characteristic of the A-form DNA as presented in sec. 3.1.1. This is consistent with the X-ray measurements presented in previous sections and with the literature (sec. 4.1 and [61]) which proved that the ethanol-induced B-to-A transition happens at ethanol concentrations between 60%-70%. It is important to notice that the 10^{th} layer peak still remains although it is much weaker. The superposition of B- and A-form features indicates that the sample is a mixture of conformations. Most of the molecules are in A-form but some of them are still hydrated enough to be in B-form. It can be said that the sample has "B-form contamination". As presented in sec. 3.1.1 evidence of conformational mixtures were previously detected using neutron scattering and other techniques.

The influence of the salt concentration on the amount of ethanol needed to cause the A-to-B transition has been previously reported [58]. Higher salt content seems to favor the A-form. Neutron data support this. Fig. 4.6a shows the high-Q part of the longitudinal reciprocal map for the Na-66%-0.01 sample. The 10^{th} layer Bragg peak is strong and the off axis peaks are weak. Fig. 4.6b show the low-Q part of the map. A Bragg peak appears at $(Q_H, Q_K) = (0.45, 0.3)$ which is usually related with A-form DNA (sec. 3.1.1). These results suggest this sample is composed predominantly by B-DNA with some A-form contamination

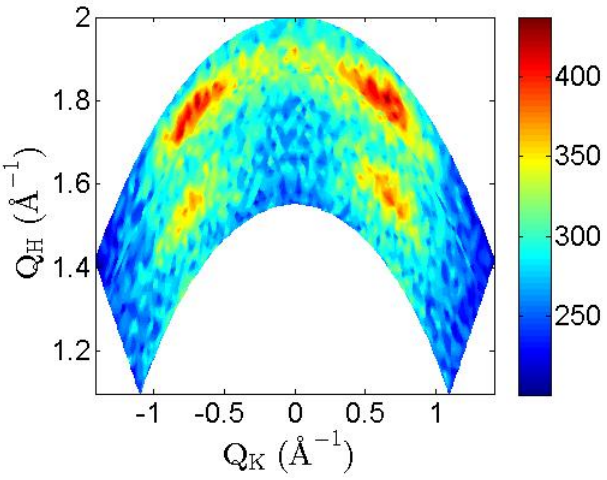


Figure 4.5: Reciprocal space map in longitudinal orientation of the Na-66%-0.1 sample. Measured in IN3.

in opposition to the similar sample with 0.1 M $NaCl$. The lower $NaCl$ content of this sample seems to increase the number of molecules which remain in the B-form.

Fig. 4.6c shows the 10^{th} layer peak for the Na-60%-0.1 sample. It suggests that this sample is in the B-form which is consistent with the previous presented X-ray data but some very weak off-axis features may be asserted to exist at $(Q_H, Q_K) = (1.6, 0.5)$ Å^{-1}and $(1.6, -0.5)$ Å^{-1}. In addition, the low-Q part of the longitudinal reciprocal space map (shown in fig. 4.6d) shows clearly the A-form Bragg peak mentioned above. It seems like the Na-60%-0.1 sample has some A-form contamination even though it was not appreciable in the X-ray images.

Similar measurements proved that Li-DNA fibers show a pure B-form at ethanol concentrations up to 80% (compare fig. 4.6e and 4.6f with sec 3.1.1).

Hereafter Li-DNA fibers submerged in ethanol will be considered to be in a pure B-form. On the contrary, Na-DNA fibers submerged in ethanol concentrations equal or above 60% ethanol in volume would be considered to have some degree of A/B mixture of conformations.

The effect of the submersion of the Na- and Li-DNA samples in ethanol on the interaxial

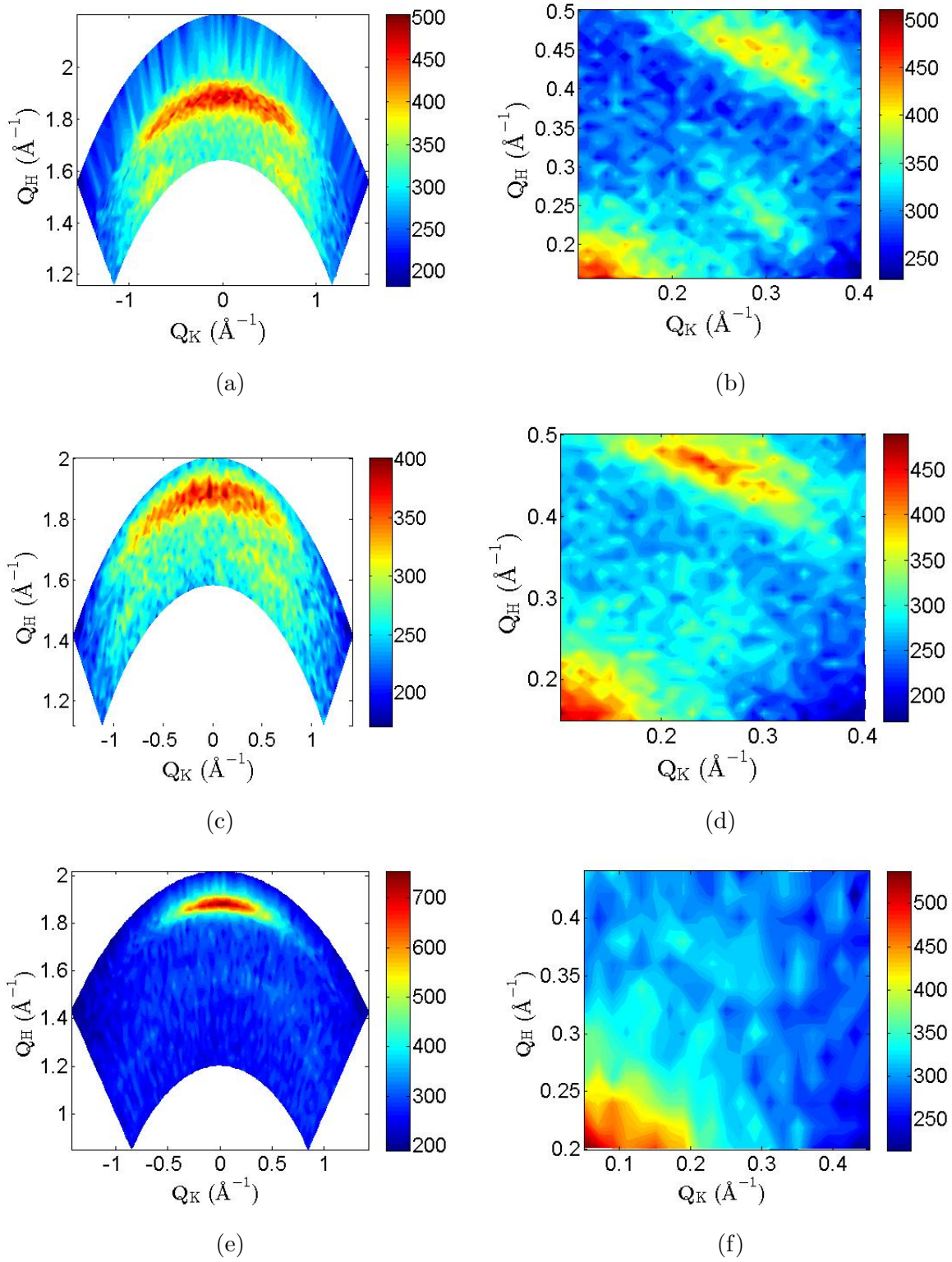


Figure 4.6: Longitudinal reciprocal space maps (left:high-Q, right:low-Q) of: Na-66%-0.01 [(a), and (b)]; Na-60%-0.1 [(c), (d)]; and Li-80% [(e) and (f)]. Measured in IN3.

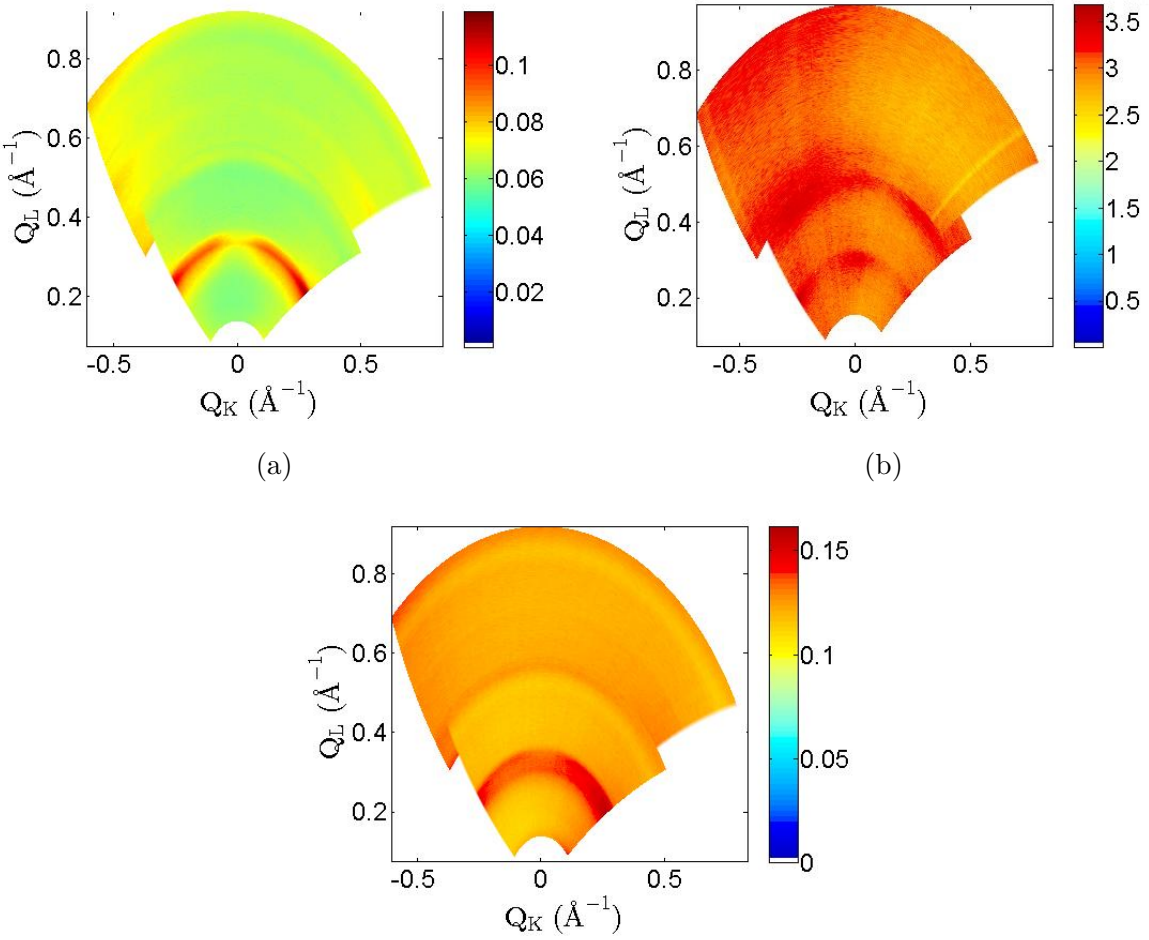
distance (I_{ad}) was addressed using the same procedure as for the PEG samples. Reciprocal space maps in the transversal orientation were measured and the intensity was summed over the sample rotation angles (for constant Q) and plotted as the magnitude of Q.

Fig. 4.7 shows representative examples of the reciprocal space maps collected on D16 to gauge the intermolecular distance of the samples submerged in ethanol solutions along with the humidified Li-fibers. Fig. 4.7a is identical to fig. 3.2a with three ring-like reflections, one strong and centered at $Q \approx 0.3$ Å^{-1} and two weaker ones at $Q > 0.3$ Å^{-1}. In the maps of the samples submerged in ethanol a feature can be seen at around 0.3 Å^{-1}which was not present in their PEG counterparts (fig. 4.4). For the Li-DNA fibers the radial distribution of this new feature is homogeneous. For the Na-DNA fibers, there is arguably a spot of higher intensity centered at approximately $(Q_L, Q_K) = (0.3, 0)$ which is similar to the peak observed in fig. 3.2b, the pattern of a pure A-form sample. This is another proof that the Na-DNA fibers submerged in ethanol have a mixture of conformations A and B.

The result of the summation over the rotation angle for equal Qs for the ethanol samples can be seen in fig. 4.8.

The data of the Li-fibers humidified at 75% RH (black in fig. 4.8a) show at least three clear reflexions at approximately 0.3, 0.58 and 0.68 Å^{-1}. The other Li-samples show corresponding reflexions. For all Li-fibers the reflexion at 0.3 Å^{-1}is composed of two peaks which are very close. This is patent for the Li-20% sample and it is suggested for the rest of the samples due to the tail of the reflexion in the low-Q part and the asymmetry of the reflexion. Therefore there are four reflexions to consider in each curve related with Li-samples. These curves are rather different from what was observed for the same measurements for PEG samples. The double-peak feature is not presented in the PEG samples and the previously observed strong peak at around 0.5 Å^{-1} appears much weaker in the ethanol samples.

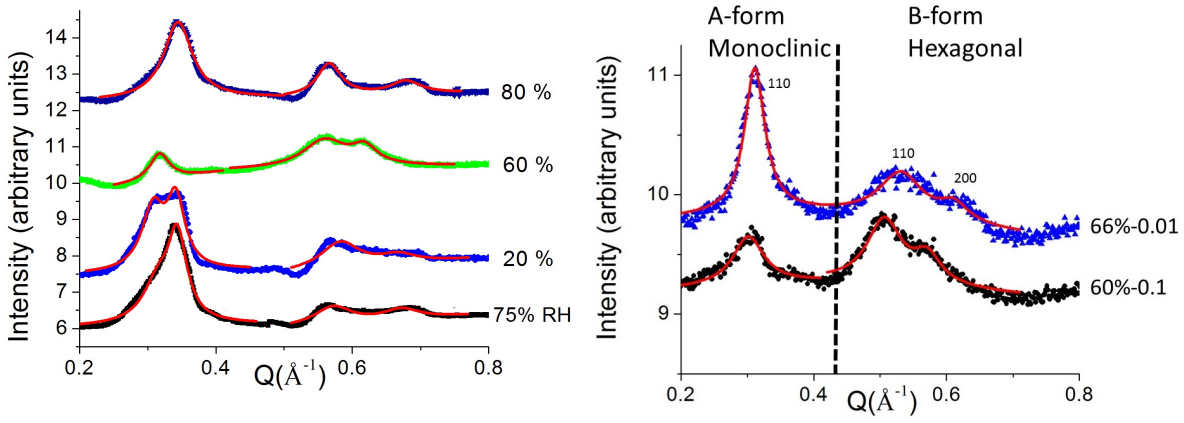
It has been reported that Li-DNA fibers humidified at 75% RH have an orthorhombic packing structure of its molecules with two molecules per unit cell in equivalent positions $(0,0,\frac{1}{6})$ and $(\frac{1}{2},\frac{1}{2},-\frac{1}{6})$ [38]. The different values in the z-axis mean that neighboring molecules are displaced with respect to each other along the fiber axis. In an orthorombic lattice the



(a) (b)

(c)

Figure 4.7: Reciprocal space maps for the tranversal orientation of several samples. a) *Li*-humid; b) *Na*-60%-0.1; and c) *Li*-20%. Measured in D16.



(a) *Li*-DNA (b) *Na*-DNA

Figure 4.8: Sum of the intensity over the sample rotation angles of the transversal reciprocal space maps of ethanol samples for constant scattering vector as a function of Q. The curves have been shifted vertically for the sake of clarity. Red lines represent the fit to the data. Panel b) includes a summary of the indexation of the *Na*-fibers taking into account their A/B conformational mixture (see text for clarification).

Sample	I_{ad} (Å)	$\frac{\Delta I_{ad}}{I_{ad,Hu}}(\%)$
Li-80%	19.40 ± 0.05	-8.96 ± 0.85
Li-60%	21.89 ± 0.04	2.72 ± 0.42
Li-20%	22.05 ± 0.02	3.47 ± 0.62
Li-humid	21.31 ± 0.06	0

Table 4.2: Percentage increase in the intermolecular spacing for Li-DNA fibers submerged in ethanol-d6 solutions with respect to the humidified Li-fibers.

four peaks seen in the data of fig. 4.8a can be indexed (from left to right) as 101, 110, 211 and 212. The relation between the d-spacing and the lattice parameters for a given hkl reflection in an orthorhombic lattice is:

$$\frac{1}{d^2} = \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \tag{4.2}$$

and the intermolecular distance (axis to axis distance) can be calculated as $I_{ad} = ((a/2)^2 + (b/2)^2)^{0.5}$.

The centers of the peaks for the data of the humidified Li-fibers were determined by fitting the reflexions with Gaussian functions over straight line backgrounds. Using the d-spacing of the four reflexions (as stated in the PEG section, $d = 2\pi/Q_C$) the intermolecular distance for the Li-DNA sample humidified to 75% RH was computed as 21.3 ± 0.06 Å which is consistent with previous works [38]. The index of the peaks were assumed to be the same for all the lithium samples submerged in ethanol and the interaxial distances were calculated analogously.

The I_{ad} for the Li-DNA fibers can be seen in table 4.2 along with the percentage increase with respect to the humidified Li-fibers. I_{ad} decreases with ethanol concentration, or alternatively whit decreasing water content in the sample which is consistent with previous works [58].

Fig. 4.8b shows the data for Na-DNA. In this case the reflexion centered at ≈ 0.3 Å^{-1}is

highly symmetric and therefore it is most probably a single peak. Therefore for the Na-DNA samples three reflexions must be taken into account. Comparing fig. 4.8b and fig. 4.8a it seems like the two reflexions centered at Q higher than 0.3 Å^{-1} differ. For the Li-DNA these reflexions are quite separated from each other (e.g. ≈ 0.13 Å^{-1} for the case of the Li-humid fibers). However, for the Na-DNA they are partially superimposed. In this sense, these reflexions look like the ones found in the PEG samples (fig. 4.4).

The indexing of the peaks for the Na-DNA samples submerged in ethanol is not as straight forward as the Li case. As commented above these samples contain a mixture of A and B conformations. Each conformation will have a different packing structure [38] and thus will give rise to different peaks. This implies that these samples will also have two different intermolecular distances, one for the packing of the A-form and another for the B form. In order to properly interpret the data of fig. 4.8b the question of which peaks belong to which conformation should be answered.

The molecules in Na-DNA fibers humidified to high RH are known to be in B-form and packed in a hexagonal lattice. When the water activity (RH) is decreased, a transition to A-form molecules packed in a monoclinic lattice takes place [38]. Increasing concentrations of ethanol decrease the water activity of the solvent (sec. 3.3.3), thus the same lattice transition will probably occur in the present samples. The samples with A/B mixed conformations are in an intermediate state in this transition. Therefore it is reasonable to assume that the part of the sample which is in A form will have a monoclinic molecular packing and the part of the sample in B form a hexagonal packing.

Fig. 4.7b shows that the Na-60%-0.1 sample has arguably a strong spot of intensity centered at approximately $(Q_L, Q_K) = (0.3, 0)$. This relates closely to the map shown in fig. 3.2b which was ascribed to pure A-form fibers. Therefore it has been assumed that the peak centered at ≈ 0.3 Å^{-1} in fig. 4.8b comes from the A-form part of the sample. This is consistent with the fact that the peak is larger in the Na-66%-0.01 sample than in the Na-60%-0.1 sample. The number of molecules in A-form in the samples will surely increase with ethanol concentration and so will do the integrated intensity of the peaks arising from

this conformation.

Taking into account the d-spacing formula for the monoclinic lattice for A-form DNA:

$$\frac{1}{d^2} = \frac{1}{\sin^2\beta}\left(\frac{h^2}{a^2} + \frac{k^2\sin^2\beta}{b^2} + \frac{l^2}{c^2} - \frac{2hl\cos\beta}{ac}\right) \tag{4.3}$$

and assuming reasonable values for a, b, c and β form the literature [38]; the peaks centered close to 0.3 Å^{-1} can be indexed as the 110 of the monoclinic lattice.

However, the d-spacing of the other two peaks in the curves of fig. 4.8b (at ≈ 0.52 and ≈ 0.56 Å^{-1} in the data of the Na-60%-0.1 sample) cannot be matched easily with eq. 4.3 and reasonable parameters. Because of this and of the similarity of these reflexions with the ones observed in the PEG samples (compare fig. 4.8b to fig. 4.4a and the higher Q reflections of the map of fig. 4.7b to the reflexions of fig. 4.2d) it has been assumed that they are related with parts of the sample which remain in B-form and has a hexagonal lattice. Once again this is consistent with the fact that these peaks are smaller in the Na-66%-0.01 sample since the percentage of B-form molecules in this sample must be smaller than in the sample with less ethanol. Therefore they can be indexed as the 110 and 200 peaks of the hexagonal lattice, as done for the PEG samples. The two peaks were fitted and the average a parameter (calculated with eq. 4.1) was taken as the I_{ad} of the sample. The results along with the percentage increase with respect to the humidified Na-fibers can be seen in table 4.3.

The intermolecular distance for the A-form part cannot be precisely determined with the current set of data because only one reflexion from the monoclinic lattice is available (so a and b cannot be unambiguously computed from eq. 4.3 even if the peak is indexed). A rough estimation of I_{ad} can be done if values from the literature are used. Assuming $\beta = 97.1^\circ$ and the ratio between the lattice parameters a and b to be $a/b = 0.54684$ as found in [38], the d-spacing of the 110 reflexion can be used with eq. 4.3 to calculate the interaxial spacing of the A-form domains of the Na-fibers submerged in ethanol. The results can be seen in table 4.3.

Sample	I_{ad} (Å)	$\frac{\Delta I_{ad}}{I_{ad,Hu}}(\%)$
B-form Na-66%-0.01	23.79 ± 0.06	-3.52 ± 0.33
A-form Na-66%-0.01	≈ 24.08	≈ -2.35
B-form Na-60%-0.1	25.51 ± 0.03	3.45 ± 0.51
A-form Na-60%-0.1	≈ 24.89	≈ 0.93
Na-humid	24.66 ± 0.09	0

Table 4.3: Percentage increase in the intermolecular spacing in the B- and A-form domains of the Na-DNA fibers submerged in ethanol-d6 solutions with respect to the humidified Na-fibers.

The values given for the I_{ad} of the A-form domains are only a rough estimation but are reasonable with respect to the values of the B-domains. Even with the added difficulty of the conformational mixture, the data show that increasing the ethanol concentration decreases the intermolecular distance as expected from previous works [58] and observed in the case of the Li-fibers.

These data show that submerging the samples in ethanol solutions can indeed modify the I_{ad} and thus ethanol solutions are a viable alternative to the osmotic pressure method to study the effect of the confinement in the melting transition.

4.3 Melting transition studied by calorimetry

Differential scanning calorimetry provided a thermodynamics view of the melting transition of the submerged fibers.

4.3.1 Fibers submerged in PEG solutions

Calorimetry data were collected in Na-DNA fibers submerged in solutions of PEG of molecular weight 1000, 6000 and 8000 g/mol (PEG1K, PEG6K and PEG8K). All the samples

were prepared with an identical buffer (sec. 2.1.1). The PEG-concentration range for the studied samples was $5 < C_{\%} < 40$ where $C_{\%}$ is the PEG concentration in weight percent.

Three other types of samples were measured in addition to fibers submerged in PEG solutions: fibers equilibrated in a humid atmosphere with 92% relative humidity (Na-humid fibers); the same DNA used for the preparation of the fibers (salmon testes) dissolved in the buffer used to prepare the PEG solution; and fibers submerged in the buffer. Note that submersion in PEG-free buffer caused partial dissolution of the fibers, but due to the relatively low solubility of the DNA in water (10 mg/ml for a saline buffer according to the provider) most of the DNA remained in fiber form.

The calorimetry data for every sample showed a well defined single-peak related with the excess heat capacity of the melting transition. Fig. 4.9a shows representative examples of the data. It displays the specific heat as a function of temperature for the humid fibers and fibers submerged in a solution of PEG6k-20%, along with the respective fits to the curves using a Gaussian function and the fraction of closed base pairs (f_{DS}) calculated using eq. 2.6. The melting temperatures, T_m, are highlighted.

Fig. 4.9b shows the T_m of: the humid fibers (92% RH); the DNA dissolved in the buffer; the fibers submerged in PEG-free buffer; and of the submerged fibers in PEG solution as a function of PEG concentration ([PEG]).

DNA molecules dissolved in solution are less confined than the DNA molecules of the fibers submerged in the PEG-free buffer. Except for that, the DNA of both samples are subjected to the same conditions (hydration, ionic environment etc). The T_m of DNA fibers submerged in buffer is higher than T_m for the DNA dissolved in the same buffer. Therefore it is probable that confinement is the origin of the higher T_m for the submerged fibers. A similar effect was reported previously in [99], in this study the transition between DNA dispersed in solution and an aggregated liquid-crystalline state increased the T_m measured by differential scanning calorimetry. More recently, consistent results were published in [100] where a shift of the DNA melting temperature of several degrees occurred due to the aggregation of DNA molecules. As explained in sec. 3.3.1, interaction between densely packed DNA molecules

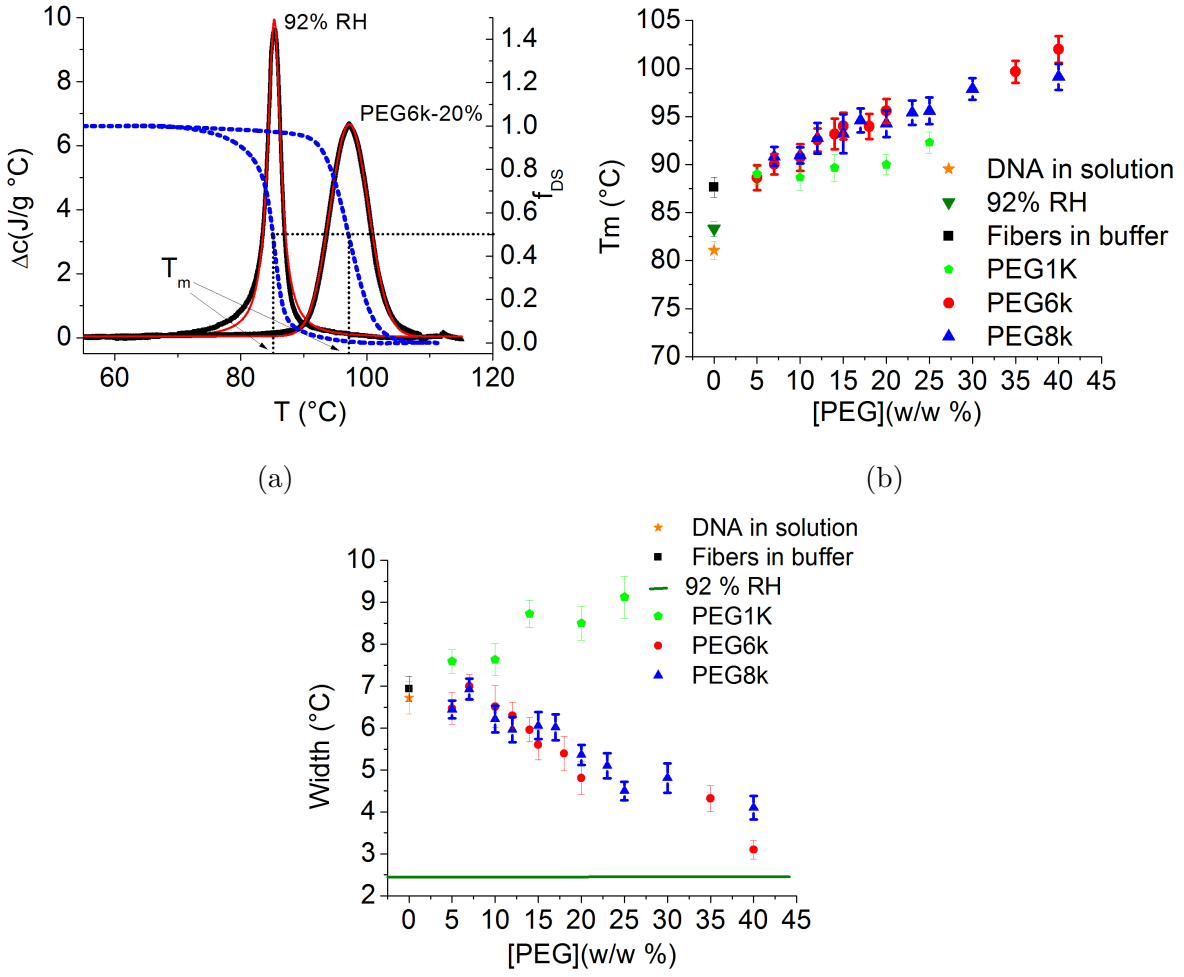


Figure 4.9: a) The heat capacity per unit mass, Δc, for *Na*-DNA fibers equilibrated in an atmosphere and 92% RH and submerged in a solution of PEG6k-20% w/w is plotted with solid thick black lines, the solid thin red lines represent the fit to the data, the blue dashed lines are the fraction of closed base pairs as a function of temperature. The T_ms are highlighted. b) Melting temperature measured of a dilute solution of DNA, DNA fibers humidified at 92% RH, fibers submerged in buffer and fibers submerged in solutions of PEG1k, PEG6k and PEG8k as a function of the [PEG]. c) Width of the transition for the same samples of panel b). The green line marks the value of the width for the humid fibers. Error bars represent the standard deviation of measurements on identical samples.

can theoretically increase T_m.

The T_m for PEG samples were consistently higher than for fibers submerged in buffer with no PEG. T_m increases linearly with [PEG] and the trend is equivalent (within error bars) for PEG6K and PEG8K. The behavior of T_m as PEG1K concentration increases is appreciably different.

Similar effects were reported for DNA dissolved in PEG solutions [82, 83] and were explained via the excluded volume effect (sec. 3.3.2). The excluded volume could be the main effect explaining the stabilization caused by PEG1K. PEG1K is significantly shorter than PEG6K which is supposedly the shorter PEG which is excluded from the DNA fibers ([67] and sec. 3.1.3). If the PEG1K can penetrate the fibers it will share the same space as the DNA and the sample will be qualitatively similar to a very dense DNA solution in which the excluded volume would be an important factor.

However, for PEG6K and PEG8K it is reasonable to assume, based on previous evidence, that the PEG molecules did not penetrate the fibers [11] and thus only a small percentage of the DNA molecules in the sample are in contact with PEG molecules. Knowing that the width of the macroscopic DNA fibers is $\approx 0.5\ mm$, if one assumes that the DNA molecules forming the fiber have a hard diameter of 20 Å and they are closely packed so that there is no separation between molecules, the percentage of DNA molecules in the surface of the fibers can be roughly estimated as less than 1%. Therefore the excluded volume and specific interactions of the PEG with DNA cannot affect significantly the melting transition. Moreover, the trend of T_m as [PEG] is increased seems to be the same for PEG6K and PEG8K which is consistent with the excluded volume having a insignificant effect in the transition ([82] and sec. 3.3.2).

When DNA fibers are submerged in PEG6K and PEG8K solutions, there is a well defined average interaxial distance which decreases as [PEG] increases (due to the osmotic pressure of the solution which grows with [PEG]) as can be seen in table 4.1 and also in [10, 11]. Therefore the confinement in these samples is increasing with [PEG]. As presented in sec. 3.3.1 it may be the main factor in the increase of T_m of the samples submerged in high

Parameter	Value
κ^{-1}	9.63 Å
R_a	9 Å
P	34 Å
Φ_s	0.8π
a	3.4 Å
θ_f	0.8
ϵ_s	80
f_c	0.4
σ_0	16.8 $\mu C/cm^2$

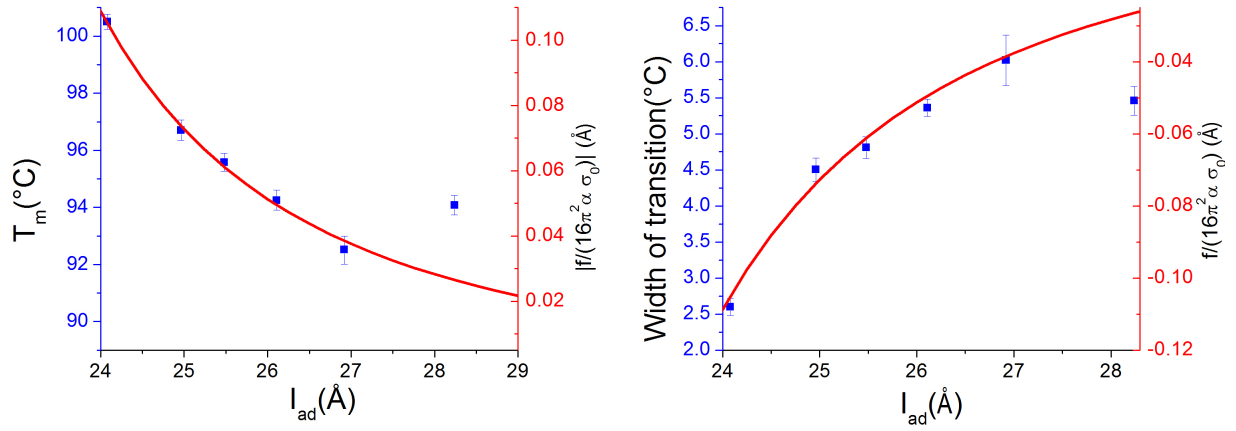
Table 4.4: Parameters used in the calculation of $f/(16\pi^2\alpha\sigma_0)$ and the theoretical melting temperature. Note that the value of κ^-1 here was used only for the calculation of the intermolecular foce. It was adjusted for the calculation of T_m (see text).

molecular weight PEG solutions.

This idea can be further explored by combining the calorimetry data for T_m with the interaxial distance (I_{ad}) measured by neutrons and the models of DNA confinement developed by Kornyshev and Leikin presented in sec. 3.2. Fig. 4.10b shows a comparison of this model with our experimental data. The magnitude $f/(16\pi^2\alpha\sigma_0)$ is proportional to the interactive force between two DNA molecules as predicted by the model [68]; calculated using equations 3.6. The parameters used are presented in table 4.4. (They are R_a, P, Φ_s and κ^-1 or Debye length). The values for R_a and P are reasonable for B-form DNA, 0.8π is a usually accepted value from the literature [68] and the Debye lenght was calculated based on the salt content of the solution (0.1 M $NaCl$) as done in [101]. The model predicts an attractive force between molecules which increases as the interaxial distance decreases.

Fig. 4.10 shows that both T_m and the width of the transition are quite correlated with the interaction force over most of the range studied which supports the idea of the confinement being the main factor in the effect of the PEG in the fiber samples.

This correlation does not hold at $I_{ad} = 28.24$ Å. Within the model the attraction between



(a) (b)

Figure 4.10: a) Left scale: Blue squares are the melting temperatures measured by calorimetry as a function of the interaxial distance between molecules (I_{ad}). Right scale: the red line is proportional to the absolute value of the attractive force between molecules predicted by eq. 3.6. b) Left axis: Blue squares are the width of the melting transition measured by calorimetry as a function of I_{ad}. Right axis: the red line is proportional to the force of the previous panel.

DNA molecules is closely related to the non-random alignment of opposite strands which is defined by the relative rotation of the two helices (Φ_1-Φ_2) [68]. The selection of Φ_1-Φ_2 in the present calculation was based on a minimization of the interaction energy which holds at distances shorter than R^*, the critical separation, given by eq. 3.12. Substituting the parameters used for computing the force of fig. 4.10 into eq. 3.12 gives $R^* \approx 29.32$ Å. This could certainly explained the breaking of correlation between experimental data and model at $I_{ad} = 28.24$ Å.

Future works will try to improve the matching of model and experimental data by changing the approximation of Φ_1-Φ_2 as R^* is approached.

In addition to this qualitative comparison between theory and experiment, a theoretical prediction of the increase of T_m as I_{ad} decreases was tried using eq. 3.16. The parameters used for this calculation were taken from the literature [69] and are presented in table 4.4 except for κ^{-1}. θ_f is the fraction od DNA charge compensated by adsorbed cations, f_c is the fraction of adsorbed cations in the minor groove of DNA and ϵ_s is the dielectric constant of the solution (for which the value of water was taken).

A Debye screening length of 9.63 Å is a reasonable value for the salt concentration of the solutions the samples were submerged in. However, when used in eq. 3.16 it gave a very poor resemblance with the data. Fig. 4.11 shows a comparison of the theoretical and experimental T_m as a function of I_{ad} for a calculation with $\kappa^{-1} = 88$ Å. Eq. 3.16 allows to predict the variation of T_m when I_{ad} changes with respect to a reference T_m, which is highlighted in the figure, and for which the theoretical T_m is assumed to match exactly the experimental data.

Fig. 4.11 shows that, with this large value for κ^{-1} the model matches the data remarkably well. Theory and model differ for the T_m at 28.24 Å, this prediction of T_m is based in the calculation of the force presented in fig. 4.10 (sec. 3.3.1 and [69]) so the range of applicability should be the same. Therefore it is not surprising that the prediction of T_m fails at $I_{ad} = 28.24$ Å.

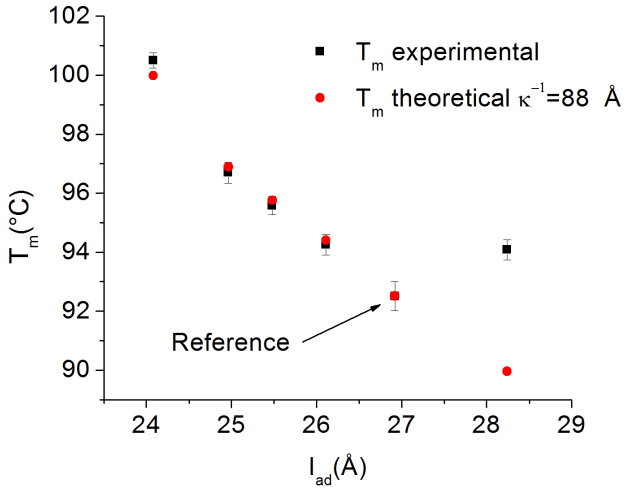


Figure 4.11: Comparison between experimental and theoratical T_m as a function of I_{ad} for a Debye screening length of 88 Å.

$\kappa^{-1} = 88$ Å relates to a solution with a salt concentration roughly two orders of magnitude lower that the expected Debye length for the solutions of the samples. The samples could exhibit a smaller ion activity than expected because of the presence of PEG which was not accounted for in the calculation of κ. However, a two orders of magnitude decrease in the ionic strength looks unreasonable. The model, which gets the T_m right just by changing one parameter, may have a relatively small shortcoming.

4.3.2 Fibers submerged in ethanol

Calorimetry data were collected in Na/Li-DNA fibers submerged in ethanol solutions as a function of the ethanol concentrations between 20% and 85% (v/v). The solutions are described in sec. 2.1.1. All samples for the calorimetry study of this section contained 0.1 M of the corresponding salt.

Fig. 4.12a shows examples of calorimetric curves collected in Na- and Li-DNA fibers submerged in ethanol solutions. The curves of Li-DNA fibers in the whole concentration range studied were qualitative identical to the PEG case with a single peak on a constant

background as seen in the figure for Li-60% ethanol (black squares). This is also the case for Na-DNA for ethanol concentrations up to 50%. However, at 60% ethanol (red circles) a tail appears at the high temperature side of the peak making it appreciably asymmetric and hinting at the existence of a second, weaker, contribution to the melting curve apart from the main strong peak. For Na-66% fibers (blue triangles in the figure) this second contribution to the melting curve is patent since there is a smaller melting peak at higher T with respect to the strong main one.

Na-DNA fibers showed some degree of A-form contamination when submerged in 60% or higher ethanol solutions (sec. 3.1.1 and sec. 4.2.2) while Li-DNA remained in B-form for the whole ethanol concentration range studied. These changes in the shapes of the melting curves for Na-DNA with increasing ethanol concentration are most probably related with the appearance of A-contamination and the increase of the number of molecules in A-form as the ethanol concentration is increased. Within this interpretation, it is reasonable to assume that most of the molecules in the Na-66%-0.01 were in B-form and underwent the melting transition with a melting temperature $\approx$ 63 °C. The fraction of the sample in A form remained in the helical form and melted at $\approx$ 70 °C(the center of the secondary peak).

Two different melting temperatures could be unambiguously computed for curves exhibiting two clear and separated melting peaks. This is not the case in the samples of the present study. All the secondary peaks ascribed to the melting of the A-form domains are relatively very small and superimposed with the main peak. A melting temperature for A and B domains could still be roughly estimated by fitting two peaks to the curve. However, even for single peak curves, identifying the melting temperature with the center of the peak can introduce errors in the calculation [102]. In this work a singular T_m (calculated as explained in sec. 2.2.5) was ascribed to the double-peak samples. These values should be understood as weighted averages between the T_m of domains in B and A conformations for each sample. Special care must be taken if the T_m values reported here are to be analyzed quantitatively. For the qualitative discussion presented in this chapter the values are probably sufficient.

On the contrary, calculating the width of the double-peak samples as the full width at

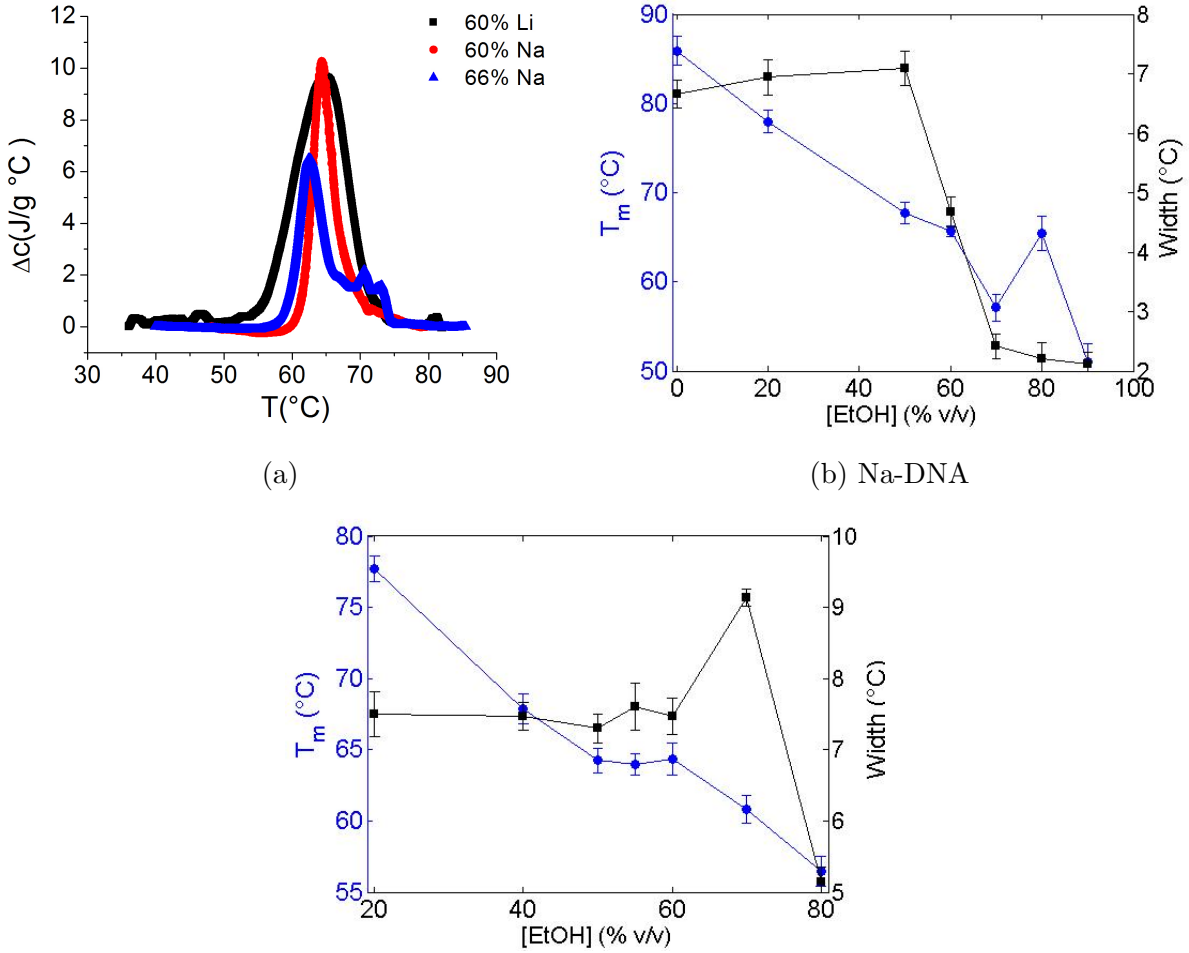
half maximum of a single peak (as described in sec. 2.2.5) would be quite meaningless. For all the Na-DNA fibers submerged in ethanol the width of the transition was calculated as the temperature change necessary to go from 25% to 75% open base pairs(using eq. 2.5). For a single-peak melting curve this is equivalent to take the full width at half maximum of the peak. This procedure was used before, with calorimetry and other techniques able to access the fraction of open base pairs, to compute the width of the transition [103].

Fig. 4.12b shows the value of the melting temperature and the width of the transition for the Na-DNA case. The melting temperature decreases from 0% until 70% ethanol and raises from 70% to 80% before decreasing again at 90%. Caution must be applied when taking into consideration the increase in T_m at 80% since only one point constitutes such increase. However, as explained previously (sec. 2.2.4) the presented values of T_m are the average of the measurement on three identical samples. Therefore we believe this increase is significant. The width of the transition is roughly constant until at least 50% ethanol and drops for higher concentrations.

Fig. 4.12c shows the same data for Li-DNA fibers, in this case T_m decreases monotonically (the local maximum in T_m does not exist) and the width increases significantly at 70% ethanol and then drops steeply at 80%.

For the ethanol samples the main factors explaining the overall effect of the ethanol in the melting temperature is decreasing of the water activity and the dielectric constant of the solution as presented in sec. 3.3.3.

The behavior of T_m and the width of the transition as a function of the ethanol concentration found in this work are qualitatively identical to what was reported by Rupprecht et al. (sec. 3.3.3) for very similar samples. The quantitative difference can be explained by differences in salt concentration of the solutions which affect both T_m and width [103]. Rupprecht et al. also found a B-to-A transition in Na-DNA (consistent with the scattering data of previous sections). They even suggested that a range of ethanol concentrations exist in which the fibers are a mixture of B and A form (fig. 3.4a and [88]) which relates to the A/B contamination found in the samples of this work.



(a)

(b) Na-DNA

(c) Li-DNA

Figure 4.12: a) Examples of calorimetric curves of $Na-$ and $Li-$DNA fibers submerged in ethanol solutions. Melting temperature and width of the transition for fibers submerged in ethanol solutions as a function of the ethanol concentration (v/v).

Rupprecht et al. noticed that, for Na-DNA fibers, the increase in T_m which leads to the local maximum and the steeply decrease of the width start at roughly the same ethanol concentration than the A-to-B conformational transition. They suggested that the local maximum in T_m and the drop of the transition width could be related with the stronger interactions that the A-form exhibits with respect to the B-form in aggregated DNA [88].

This seems to be also true for the Na-fibers samples of the present work. In fig. 4.12b the width starts to decrease at 60% ethanol and stabilizes close to its minimum at 66%. The T_m starts to increases at 66% and both fig. 4.12a and the neutron data (sec. 4.2.2) suggest that the B-to-A transition starts at some point around 60%.

Therefore the results of fig. 4.12b can reinforce the hypothesis that the intermolecular interactions affect the parameters of the transition. It is likely that as the intermolecular interactions increase, the T_m increases and the width of the transition decreases. It is important to keep in mind that intermolecular interactions are not the only thing that the B-to-A transition will change, the stacking of the bases and even the heat of melting of base pairs in the A-form parts will be different from the B-part due to the significant different geometry of the molecule.

Summarizing, the non monotonic behavior for T_m and the width of the transition observed in Na-DNA fibers as the ethanol concentration is raised is most probably related with the B-to-A transition but no single mechanism can explain how this transition affects the melting curves.

The drop of the width of the transition for the Li-DNA fibers (fig. 4.12c) is possibly related to the B-to-C transition which was also detected by Rupprecht et al. in these samples (sec. 3.3.3). Why this conformational transition affects only the width and not the T_m of Li-DNA fibers is still an open question.

4.4 Study of the melting transition with neutron scattering

Reciprocal space maps in the (Q_H, Q_K) plane were collected at different temperatures (T) for several PEG and ethanol samples to follow the evolution of the Bragg peak at $Q_H \approx 1.87$ Å^{-1}(10^{th} layer Bragg peak). Data along $\vec{Q_H}$ when $\vec{Q_K} = 0$ were extracted for each temperature. These cuts through the maps go through the center of the Bragg peak.

The integrated intensity of the peak is directly proportional to the number of closed base pairs in the intact double helix domains. As the temperature rises and the base pairs dissociate, the integrated intensity of the peak is expected to decrease. The width of the Bragg peak is inversely proportional to the average size of the intact double helical domains. It gives access to the evolution of the correlation length of the molecule as a function of temperature.

4.4.1 Fibers submerged in PEG

The DNA-PEG samples investigated as a function of temperature were: PEG6k-17%, PEG6k-20%, PEG8k-15% and PEG8k-20%. Two cassettes containing PEG-only solutions were prepared. The cassettes contained PEG6k-17% and PEG6k-20% respectively. The temperature-dependent scattering of the PEG-only samples was measured to account for the contribution of the PEG solution to the scattering of the sample.

The cuts along $\vec{Q_H}$ at $\vec{Q_K} = \vec{Q_L} = 0$ in the maps of the DNA samples for every temperature were concatenated in order to create fig. 4.13. Each scan showed a mixture of Bragg and diffuse scattering. The Bragg scattering consisted of the DNA Bragg peak at $\vec{Q_H} \approx 1.87$ Å^{-1} and other Bragg peaks related with the material of the sample cassette. The diffuse contribution gave rise to a nonzero Q-independent background plus a broad bump underneath the Bragg peaks. The DNA Bragg peaks are relatively unchanged until high temperatures are reached, where they disappear. Every sample was cooled back to room temperature after the

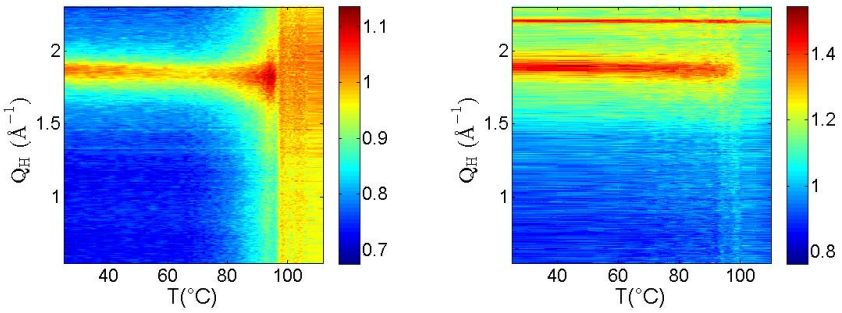
measurement at the highest T and then measured again. The Bragg peak was not recovered in any sample, showing that the DNA had completely melted at high T and the orientation given by the fiber structure had been lost. This did not happen in the pilot study [12]. In this work the heating was stopped before the melting transition was completed, and a Bragg peak was recovered when the sample was cooled back to room temperature [12].

The PEG6k-17% sample was measured on D19 and the data are shown in fig. 4.13a, the figure shows that there is an increase of the diffuse background as the transition is approached, particularly for $T \geq 97$ °C. This effect must be related with the DNA since the scattering from the PEG solution alone did not change significantly with T.

Figures 4.13b, 4.13c and 4.13d show data from samples measured on WOMBAT. An increase of the diffuse scattering with temperature is also present in these samples but it is weaker and barely noticeable in these figures. The narrow peak appearing at ≈ 2.2 Å^{-1} in the WOMBAT data was attributed to a Bragg peak from the lead seal of the cassette. It does not appear in the D19 data due to the smaller footprint of the beam. The width of the beam on WOMBAT (25 mm) illuminated the lead seal of the cassette but on D19 the circular beam of diameter 10 mm did not.

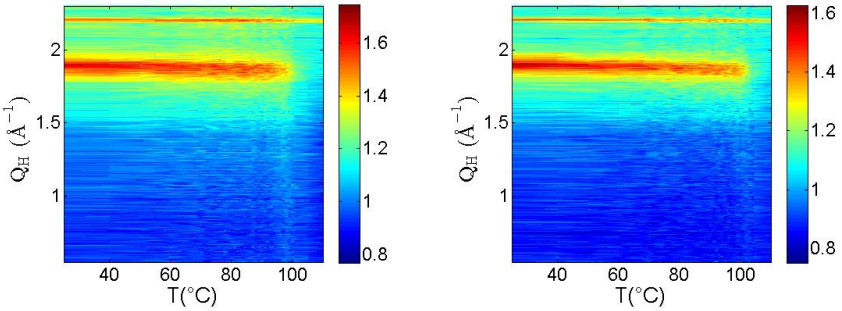
Representative data for the $\vec{Q_H}$ scans of the PEG6k-17% sample measured on D19 and for the PEG6k-20% sample measured on WOMBAT at different T are shown in fig. 4.14a and fig. 4.14b respectively. The DNA Bragg peak at $\vec{Q_H} \approx 1.87$ Å^{-1} is clearly visible. The figure includes a similar scan for the cassettes containing only the PEG solutions at room temperature. It shows that the diffraction from the PEG generates a broad bump in the same $\vec{Q_H}$ range as the DNA peak. The bump cannot be easily subtracted from the DNA curves due to the uncertainty in the liquid volumes of the two samples, meaning that the two data sets could not be accurately normalized on the same scale. The narrow peaks appearing at $\vec{Q_H} \approx 2.6$ and 3.1 Å^{-1} in the D19 data are due to Bragg peaks from the aluminum of the cassette. The Q-range recorded on WOMBAT is smaller due to the longer wavelength used with respect to D19.

The 10^{th} Bragg peak was fitted at every temperature to determine the integrated intensity



(a) PEG6K-17%

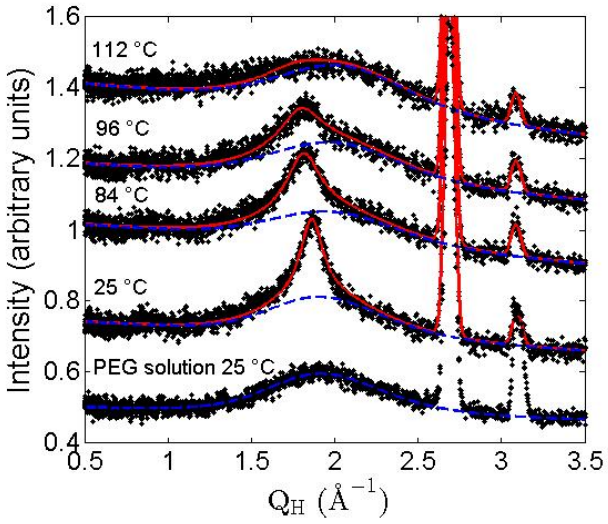
(b) PEG6K-20%



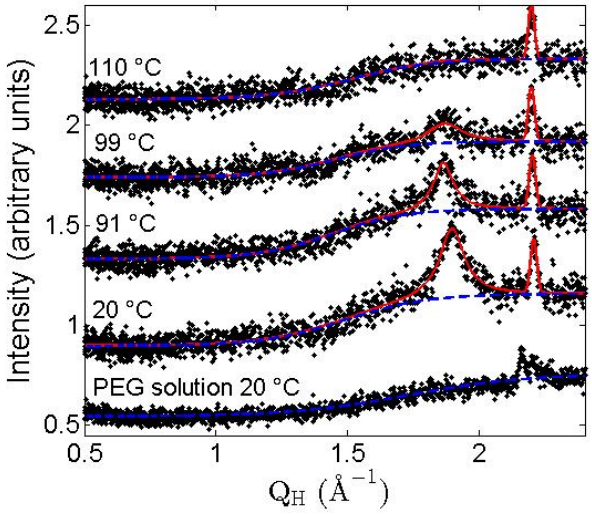
(c) PEG8K-15%

(d) PEG8K-20%

Figure 4.13: Temperature dependence of the $Q_H \approx 1.87$ Å^{-1} Bragg peak. The PEG6k-17% sample a) was measured on D19; the other samples b)-d) on WOMBAT.



(a)



(b)

Figure 4.14: Representative examples of the temperature dependence of the $Q_H \approx 1.87$ Å^{-1} Bragg peak and the PEG solution without DNA for a) the PEG6k-17% sample measured at D19 and b) the PEG6k-20% measured at WOMBAT. The red solid line is the final fit to the data using eq. 4.7, the blue dashed line is the fit of the diffuse scattering (eq. 4.5) . The curves have been shifted vertically for the sake of clarity.

and width. In previous works performed on humidified fibers [7] the peak was succesfully fitted with a Lorentzian function on a constant background. A Lorentzian is still appropriate for the Bragg peak of the present samples but the addition of the PEG solution to the DNA sample gave rise to added diffuse scattering that complicated the fitting of the peak. The origin of the diffuse scattering is not totally understood at the moment. The contribution of the PEG is important as shown in fig. 4.14 but the DNA is also contributing since it is responsible for the part of the diffuse scattering that increases with temperature.

A Lorentzian function was used to describe the DNA Bragg peak and the diffuse scattering was modeled using a phenomenological approach. The Bragg peak was modeled with:

$$I_1(Q_H) = \frac{I_0}{\pi} \frac{\gamma/2}{(Q_H - Q_0)^2 + (\gamma/2)^2} \tag{4.4}$$

where I_0 is the integrated intensity of the peak, γ is the full width at half maximum of the peak and Q_0 is the peak center. The best results for fitting the diffuse scattering were achieved using a "double-step" function on a straight line background with the form:

$$I_2(Q_H) = A_0 + BQ_H + \frac{A_1}{2}\Big[\tanh\Big(\frac{Q_H - Q_1}{s_1}\Big) - \tanh\Big(\frac{Q_H - Q_2}{s_2}\Big)\Big] \tag{4.5}$$

A_0 and B are respectively the intercept and slope of the straight line. The third and fourth terms in eq. 4.5 generate a function with a rising step centered at Q_1 and a declining step centered at Q_2 with a plateau in between whose height is set by A_1. The gradients of the first and second step are defined as s_1 and s_2 respectively. Lastly the Bragg peaks not related with DNA (Al and Pb) were accounted for with Gaussian functions:

$$I_{Al,Pb} = \sum_{i=Al,Pb} \frac{A_i}{s_i\sqrt{2\pi}} e^{\frac{-(Q_H - Q_{0,i})^2}{2s_i^2}} \tag{4.6}$$

where A_i are the integrated intensities of the Gaussian functions, $Q_{0,i}$ are their centers and s_i are their variances.

The total intensity for every temperature was fitted with the expression:

$$I(Q_H) = I_1 + I_2 + I_{Al,Pb} \tag{4.7}$$

A_0 and B were not fit-parameters. Their values were calculated before fitting a given curve. For the D19 data the average values of the scattering signal for $Q_H < 0.6$ Å^{-1}and $Q_H > 3.4$ Å^{-1} were calculated and a line was drawn between the two values. A_0 and B were taken as respectively the intercept and slope of that line. For the WOMBAT data, B was set to zero and A_0 was determined by averaging the signal for $Q_H < 0.8$ Å^{-1}. In addition, for the WOMBAT data the diffuse scattering was represented by a single rising step function so the term $\tanh(\frac{Q_H - Q_2}{s_2})$ in eq. 4.5 related with the declining step was not included.

The curves of the PEG solutions at room temperature were fitted using eq. 4.5. The results can be seen in fig. 4.14a and 4.14b. The values for the parameters of this fit (A_1, Q_1, s_1, Q_2, s_2) were used as part of the initial set of values for fitting the DNA data at room temperature using eq. 4.7. The rest of initial values were chosen by visual comparison.

For subsequent temperatures a two-stage process was used in order to get an accurate fit when the size of the Bragg peak and the diffuse contribution became comparable. For a given temperature above room T, the γ from the fit of the previous temperature was used to define a range in Q_H centered around the DNA Bragg peak ($Q_0 - 0.8\gamma \leq Q \leq Q_0 + 0.8\gamma$). In the first stage the parameters of the DNA Bragg peak (eq. 4.4) were fixed, data inside the previously defined range were ignored and eq. 4.7 was fitted using the results for the fit of the previous temperature as initial values for the parameters . In the second stage data outside the range were ignored, the parameters of eq. 4.5 and 4.6 were fixed to the fitted values of the first stage and the DNA Bragg peak was fitted with eq. 4.7. The two-stage process was then repeated twice per temperature. Each new iteration used the parameters of the previous iteration as initial parameters for the fit. After 3 iterations the parameters were stable.

Examples of the final fits can be seen in fig. 4.14. The fitting is very consistent but should be interpreted with care at high temperatures, where the Lorentzian peak was so small in

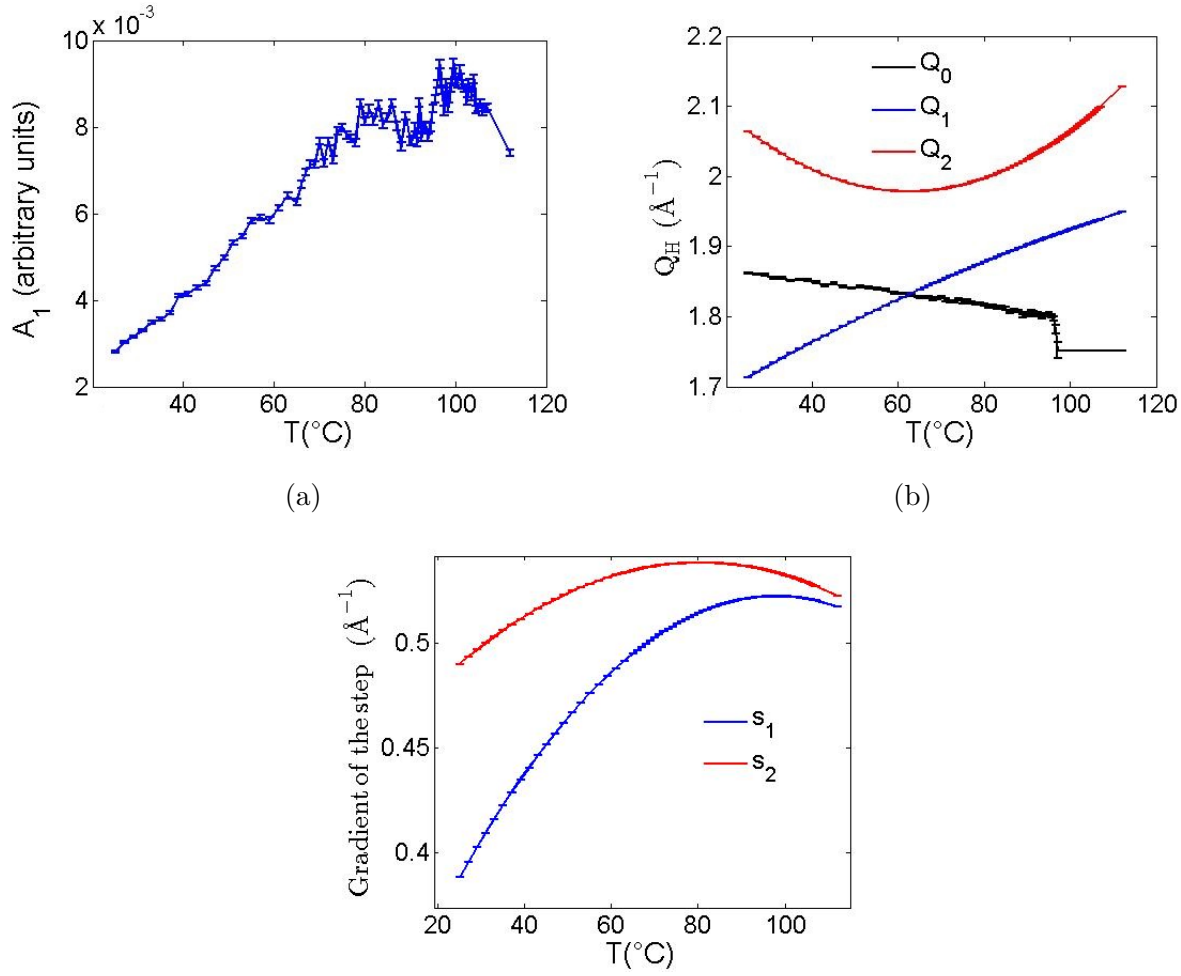
comparison with the diffuse contribution that the general fit had trouble converging. At these temperatures the values for γ and Q_0 were fixed to the values of the last temperature in which the fit converged and only I_0 was fitted.

Fig. 4.15 shows a representative example of the temperature dependence of the parameters of the double-step function and of the center of the DNA Bragg peak (Q_0). The center of the Bragg peak shifts slightly to smaller Qs as the temperature raises (fig. 4.15b) which is consistent with thermal expansion and the change in the average base pair distance due to the relaxation of the base stacking because of partial unwinding of the helix. Q_0 suddenly drops at $T \approx 97$ °C. This steep decrease could be associated to a partial untwist of the helix due to the great number of unstacked bases at this high temperature. With a sudden untwist the average distance between adjacent base pairs will grow fast and so the center of the Bragg peak will decrease. However, at this high temperature the peak is very small in comparison to the diffuse background therefore, as already mentioned, the interpretation of Q_0 and other fitted parameters must be done with care. Regarding the the parameters of the double-step, their evolution with T is reasonable but it is very challenging to ascribe a physical meaning to them.

The fitted values for eq. 4.6 (related with the aluminum and lead Bragg peaks) did not change within reasonable limits with T which is consistent with the fact that no significant changes were expected from the Al or Pb in the range of temperatures studied.

Fig. 4.16 shows the evolution of the A_0 parameter with temperature for all the samples. $A_0(T)$ in this figure is normalized by its value at $T = 20$ °C to allow easy comparison between samples. $A_0(T)$ describes adequately the magnitude of the increase of the diffuse scattering with T. The figure proves that the diffuse contribution rises close to the melting temperature in every sample. The increase at the highest T reached is approximately 20% for the PEG6K-17% (measured at D19) and approximately 5% for the rest of the samples (measured at WOMBAT) . It is likely that this quantitative difference is due to the different instrument configurations. There are two factors likely to contribute to the difference:

The first concerns the collimation between sample and analyser. D19 had no collimation



(a) (b)

(c)

Figure 4.15: Temperature dependence of the parameters of the double-step function and the center of the DNA Bragg peak for the PEG6K-17% sample as found by the fit. a) A_1, b) Centers of the double step function (Q_1 and Q_2) and of the Bragg peak (Q_0). c) Gradients of the double step function (s_1 and s_2).

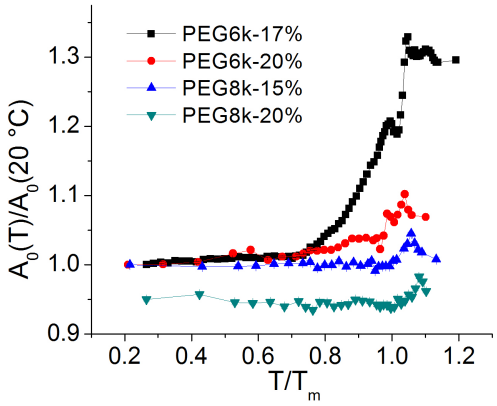


Figure 4.16: Parameter A_0 from equation 4.5 as a function of the reduced temperature (T/Tm calculated in Celsius) normalized by its value at 20 °C for each sample. A_0 is the average of the intensity at low Q and represents the magnitude of the diffuse contribution. The curve of the PEG8k-20% sample is shifted downwards by 0.05 units for clarity.

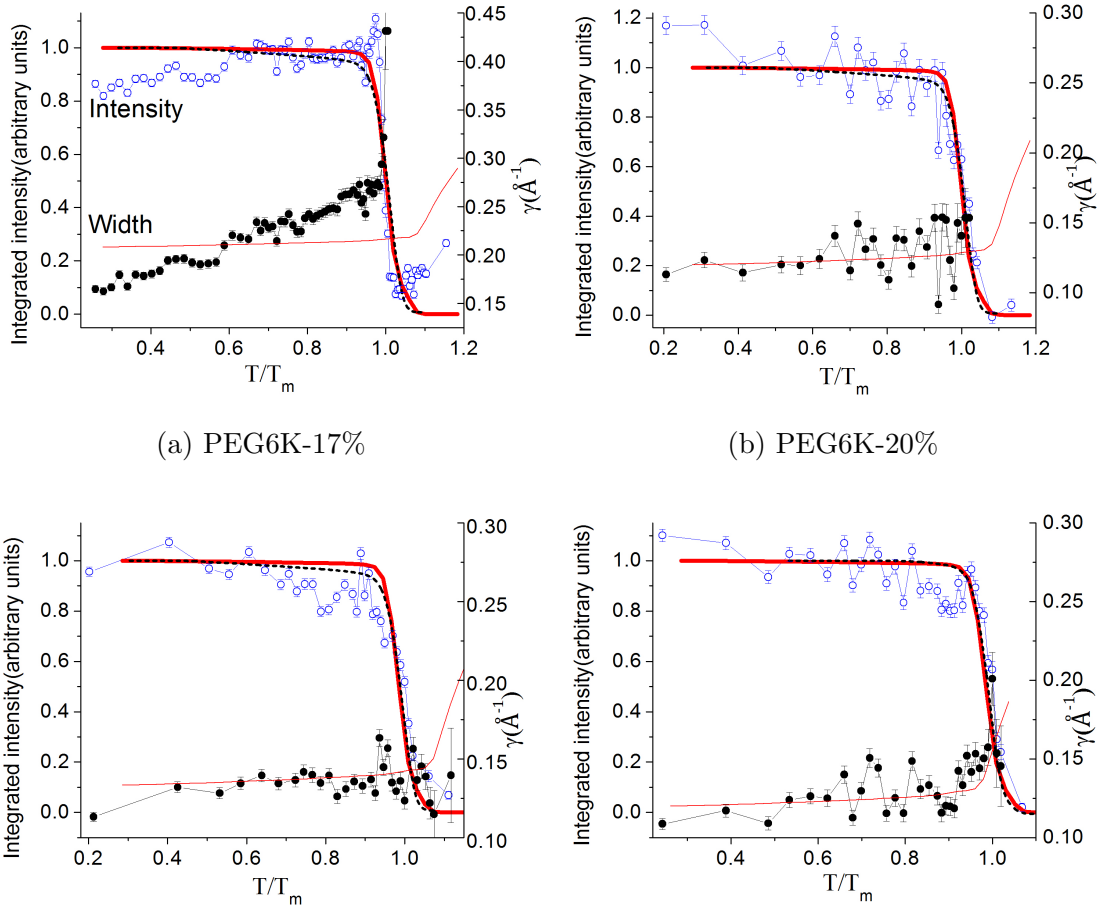
while WOMBAT used an oscillating collimator. In principle, an oscillating collimator limits only the scattering gauge volume of the sample. However, the diffuse scattering resembles, to first order, incoherent scattering while the scattering angles for the Bragg peaks are reasonably well self-collimated. The WOMBAT oscillating collimator also limits instantaneously the divergence of the scattered beam to 0.5°, hence reducing an incoherent contribution relative to a coherent signal. This partly explains the larger diffuse signal in the D19 data.

The second concerns the kinematic limits on the energy integration of the neutron instrument. Neither WOMBAT nor D19 had any means to analyze the final energy of the neutron. Both instruments therefore integrated over neutron energy transfers. The allowed energy and momentum transfers are limited by the energy of the incident neutrons, which was 14.20 meV for D19 and 3.83 meV for WOMBAT. D19 therefore integrated over a much larger energy window than WOMBAT. Fiber DNA has substantial spectral weight from 2 to 25 meV [104] , and D19 will capture these lattice dynamics where WOMBAT cannot. The dynamics of DNA will change substantially as the melting temperature is approached, giving a non-linear increase in the spectral weight for all Q and proportionately more diffuse scattering in the D19 data.

Fig. 4.16 also shows that the values for A_0 do not increase monotonically with temperature close to melting, particularly for the PEG6k-17% (black squares) and the PEG6k-20% (red circles) samples. The data shows small peaks in A_0 which may also be observed in fig. 4.13a and fig. 4.13b as vertical lines close to the temperature at which the Bragg peak disappears. As previously stated, the temperature stability at the melting temperatures for these samples was rather poor (sec. 2.4.1 and 2.4.1). The amplitude of the diffuse scattering increases steeply with temperature in this range. The pilot study [12] showed that the melting transition is, to some extent, reversible in these samples given that the melting is not complete. It is possible that the variations in A_0 are coupled with the temperature stability, and that the final registered values are indicative of the weighted average of the fluctuating intensity over the duration of the measurement.

Fig. 4.17 shows the fitted I_0 and γ as a function of temperature for the four samples. For all samples ,the intensities decreases rapidly and the peak effectively disappears over the range $0.95 \leq T/T_m \leq 1.05$. For the WOMBAT data, the integrated intensities and widths stay roughly unchanged until close to the melting temperature, as seen in previous studies on humidified fibers [7, 8]. This suggests that, as for humidified fibers, long segments of DNA remain until the very last stages of melting. These define the width of the Bragg peak. At the temperature where these last segments denature, the intensity of the Bragg peak is so weak that it is lost in the diffuse scattering and any change in the width cannot be unambiguously determined from a fit.

The PEG6k-17% sample measured on D19 shows a somewhat different behavior. Both I_0 and γ increase steadily before the T_m. This difference in the behavior of the fitted parameters with respect to the WOMBAT data may be related to the different integration over neutron energy transfers for both instruments. An increase in the phonon dynamics with increasing temperature can manifest itself as an increasing width in the tails of a Bragg peak. Such effect would be more dramatic in D19 due to the expanded energy integration. Although a neutron inelastic scattering measurement would be necessary to prove the hypothesis, it appears likely that the increasing I_0 and γ with temperature seen on D19 are due to the increased dynamics of the fiber DNA as melting is approached.



(a) PEG6K-17%

(b) PEG6K-20%

(c) PEG8K-15%

(d) PEG8K-20%

Figure 4.17: Comparison between theory and experiment for the data of the sample submerged in PEG. PEG6k-17% a) was measured on D19, the other samples b)-d) on WOMBAT. A reduced temperature, T/T_m (calculated in Celsius) was used where T_m is the melting transition of the sample. In each panel the three upper curves relate to the scale in the left, they are the fitted integrated intensity, I_0 (blue open circles), the theoretical integrated intensity (thick red line) and the fraction of closed base pairs calculated with calorimetry (eq. 2.6) for a corresponding sample (black dashed line), the fitted integrated intensities have been divided by arbitrary numbers for easy comparison. The lower curves relate to the scale in the right, they are γ, the fitted width of the peak (black close circles), and the theoretical width of the peak (thin red line).

Analysis using the Peyrard-Bishop-Dauxious model

The PBD model was used to make a quantitative analysis of the fit results, following a previously-reported procedure [7, 94]. The model can be used to calculated the probability that a base pair of a given sequence is closed at a given temperature. The probability of a given size distribution of closed segments can then be determined, as explained in sec. 3.4.1. The expected diffraction pattern, $S(Q)$, from the sequence can then be calculated (as presented in sec. 3.4.1) and compared to the experiments.

The parameters for the Hamiltonian the model uses to described the DNA molecule (eq. 3.17) were refined by comparing the calculated melting curve with the data from calorimetry.

A segment of the genome of Atlantic Salmon reported previously [105] (10000 bases starting at site 50000) was chosen for the calculation. In the modeling W was assumed to be the same for all combinations of adjacent base pairs, which has been shown to be reasonable for describing the melting curves of long DNA molecules [7, 92]. The DNA sequence was taken into account only for the calculation of the intra-pair potential V_j.

All values of the parameters of the Hamiltonian used to model the data, except for D_{AT} and D_{GC}, are the same as previously used in the study of humidified fibers [7]. They are shown in table 4.5. D_{AT} and D_{GC} represent the energy barrier for base pair dissociation and thus they are the main factor in setting the T_m of the theoretical melting curve. As shown in fig. 4.9b, the melting temperature when the fibers are submerged in the PEG solutions increases with PEG concentration. Thus D_{AT} and D_{GC} needed to be scaled to reproduced the melting temperature of each sample. The values for the depth of the Morse potential shown in table 4.5 ($D_{AT,95}$, $D_{GC,95}$) with the chosen sequence and parameters correspond to $T_m = 95$ °C. For a sample with a $T_m = T_{m,x}$ in Celsius the value for D_{AT} was scaled by:

$$D_{AT,x} = D_{AT,95}(\frac{T_{m,x} + 273}{95 + 273}) \tag{4.8}$$

and an analogous expression was used for the value of D_{GC}.

Parameters	Value
a	3.4 Å
y_c	1.5 Å
k	4.5 10^{-4} eV/Å^2
ρ	50
b	0.2 Å^{-1}
α_{AT}	4.2 Å^{-1}
α_{GC}	6.9 Å^{-1}
$D_{AT,95}$	0.13150 eV
$D_{GC,95}$	0.17150 eV

Table 4.5: Parameters used in the PBD model.

Sample	$<\Lambda^2>$(Å)
PEG6k-17%	0.45
PEG8k-15%	0.35
PEG8k-20%	0.33
PEG6k-20%	0.33
Humid fibers	0.18

Table 4.6: Standard deviation of the distance between consecutive base pairs used to calculated the $S(\vec{Q})$ of each sample.

Regarding the calculation of the theoretical $S(Q)$ (sec. 3.4.1); a value of $<\Lambda^2>= 0.18$ Å, reported by Lavery et al. [106] , was used to model the melting of humidified DNA in previous works [7]. This value gave widths for the Bragg peak that were too small in comparison to the experiemntal value of the submerged samples. $<\Lambda^2>$ was increased to model the current data. The values used can be found in table 4.6.

It is unlikely that the effects of the PEG can directly modify the standard deviation of the distance between base pairs along the molecule. More likely, the increase of $<\Lambda^2>$ accounted phenomenologically for other disorder effects, such as an increased variation of

the alignment of the molecules, when the fibers were submerged.

The Debye-Waller factor, Δ, is a function of both $<\sigma^2>$ and $<\Lambda^2>$ and it is possible that the increased Bragg widths are at least partly due to thermal fluctuations that were not linear with temperature. However, without inelastic scattering data it is not possible to say whether the larger widths were due to increased structural disorder, increased thermal fluctuations, or a combination of both factors.

The calculated $S(\vec{Q})$ for each T was fitted with eq. 4.4 to determine values for I_0 and γ as a function of T that could be compared to the experimental data.

Fig. 4.17 shows a comparison of the experimental and theoretical integrated intensities and widths of the peaks for each sample along with the fractions of open base pairs calculated by calorimetry. The data are plotted as a function of reduced temperature, T/T_m, as there are differences between the temperatures recorded with neutrons and calorimetry due to different sample environments.

Overall, the model matches all the data quantitatively. The agreement is excellent for the samples measured on WOMBAT. While the magnitudes for the model match the D19 data and the model does predict a linear increase in width with temperature, the model fails to adequately describe the gradient of the increasing width and of the integrated intensity with temperature below T_m. This may be an aspect of the modelling that does not fully capture the DNA phonon dynamics which, combined with the expanded energy integration on D19, are the possible cause of the increases in I_0 and γ. However, the magnitudes of the intensities and widths are matched and modelling may be regarded as satisfactory.

Discussion

The data are an improvement on the pilot study of DNA fibers submerged in PEG solutions [12]. The use of PEG solutions allowed the fibers to swell without losing their orientation. The neutron experiment was able to follow the evolution of the Bragg peak with temperature. The PBD model was able to describe the data, and the description was particularly good

for the WOMBAT data.

The previous study showed that the transition is, to an extent, reversible [12]. In this sense, it is similar to a critical phase transition. The diffuse scattering in fig. 4.16 increases rapidly at the melting temperature. The PEG6k-17% and PEG6k-20% samples show departures from a monotonic increase in the diffuse scattering intensity close to the melting temperature. These measurements suffered from relatively large temperature fluctuations around the melting temperature, and it is likely that the recorded intensity variations are a result from the time-average of the concomitantly fluctuating diffuse signal. The measurements with better temperature stability at the melting transition did not show this anomalous intensity variations.

These variations do not appear to be present in the integrated intensities of the Bragg peaks of the affected samples. A Bragg peak is due to coherent scattering from long-ranged, time-averaged order, while the diffuse scattering is more incoherent in nature. It may be that the diffuse scattering is dominated by the DNA dynamics, particularly of open base pairs, and its temperature dependence is driven by changes in the spectral weight. This hypothesis is consistent with the observation that D19, with the larger energy integration, shows a larger increase in the diffuse scattering.

The PBD model uses one set of parameters for DNA and is able to satisfactorily reproduce all the data despite the change in confinement brought about by the osmotic pressure. This shows that the model is, within reasonable limits, insensitive to confinement of the DNA and establishes that the analysis shown in previous works [7, 8, 94] is sound.

It is noteworthy that most of the parameters in table 4.5 are the same as those for the humidified fibers [7, 94] and for genomic DNA in solution [92]. The only differences lie in the values of D_{AT} and D_{GC}. Regarding the model, the unique effect of the submersion of the fibers in PEG solutions on the melting transition seems to be in adjusting the apparent dissociation energy of the base pairs.

4.4.2 Fibers submerged in ethanol

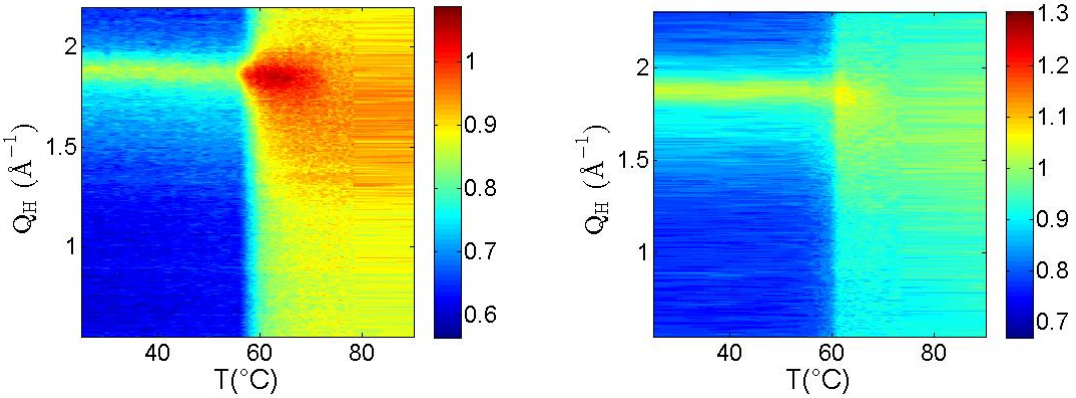
Most of the temperature dependence study was repeated for ethanol samples. The analysis of these data is still a work in progress. Here the experimental results and preliminary conclusions are reported.

Temperature-dependent scattering data of three samples were collected on D19: Na-DNA fibers submerged in 60% ethanol-d6 with 0.1 M $NaCl$ (Na-60%-0.1); Na-DNA fibers in 66% ethanol-d6 with 0.01 M $NaCl$ (Na-66%-0.01); and Li-DNA fibers in 60% ethanol-d6 with 0.1 M of $LiCl$ like the rest of the Li samples (Li-60%). A cassette containing only 60% ethanol-d6 solution was measured as a reference for the ethanol samples. The changes of the scattering of the reference cassette with temperature were negligible.

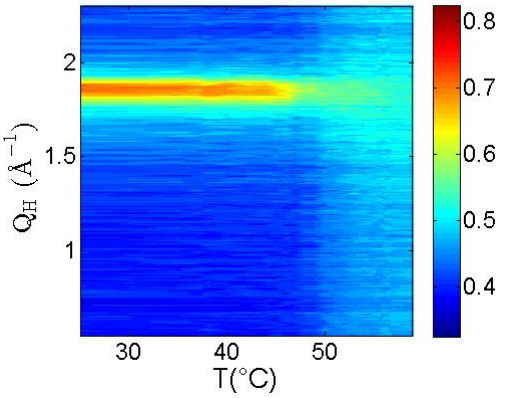
As previously done for the PEG samples, reciprocal space maps in the longitudinal orientation were collected as a function of temperature. The reciprocal space maps at every temperature for each Na sample were examined carefully for signs of the A-form (the off-axis peaks) and none were found. However, these samples are the same as studied previously in D16 and IN3 (sec. 4.2.2) and, on those instruments, both the transversal maps and the low-Q longitudinal maps proved the presence of some A-form contamination. Whether this contamination is negligible or has an impact in the results and interpretation of the temperature-dependent data for these samples is not clear at the moment.

Data along Q_H when (Q_K=0) for every temperature were extracted and concatenated in order to create fig. 4.18. Following these figures, the results are similar to the case of the PEG samples. Generally the peak-signal-to-noise ratio is significantly better due to the absence of the long protonated PEG molecules. The scans for the DNA samples still are a mixture of coherent (Bragg peak) and diffuse (incoherent plus broad bump under the Bragg peak) scattering. The peaks remain approximately unchanged until high temperatures. The diffuse scattering increased significantly for all the samples.

Fig. 4.19 shows representative examples of the scans along the axis for different temperatures for the Li-60% sample. The peak appearing at $\approx$ 2.6 Å is again related with the



(a) Na-60%-0.1 (b) Na-66%-0.01



(c) Li-60%

Figure 4.18: Temperature dependence of the Bragg peak at $Q_H \approx 1.87$ Å^{-1} for the ethanol samples. Measured at D19.

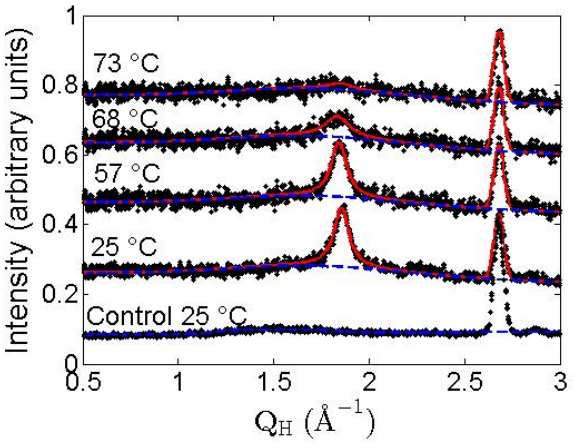


Figure 4.19: Examples of the temperature dependence of the $Q_H \approx 1.87$ Å^{-1} Bragg peak for *Li*-DNA fibers submerged 60% ethanol-d6 solution. The scattering of a cassette filled just with the ethanol solution is presented as a control. The red solid line represents the final fit to the data using eq 4.7 and the dashed blue line is the fit of the diffuse scattering using eq. 4.5.

aluminum of the sample cassette. A similar scan for the cassette containing only the ethanol solution is included in the same figure. The contribution of the ethanol solution is featureless in comparison to that of the PEG solution.

Even though the contribution of the solution does not generate a broad bump for the ethanol case as it did for the PEG. The diffuse scattering from the DNA still gives a significant contribution at the base of the Bragg peak. Therefore a simple Lorenztian on a constant background was (as in the PEG case) not able to fit the data. The data from ethanol samples were fitted with the same expression described in the previous section to fit the D19 data (eq. 4.7). The parameters of the double-step had to be fixed for every temperature for the fit to converge. This indicates that the double-step function is not as optimal as in the PEG case for describing the diffuse scattering of the ethanol samples. The parameters used for every sample are presented in table 4.7.

Fig. 4.20 shows the parameter A_1 of the double-step function (eq. 4.5) and the center of the DNA Bragg peak (Q_0) as a function of temperature from the fit of the *Li*-60%.

Sample	Q_1 (Å^{-1})	Q_2 (Å^{-1})	s_1	s_2
Na-60%	1.22	2.4	0.33	0.50
Na-66%	1.19	2.39	0.31	0.51
Li-60%	1.81	1.90	0.32	0.49

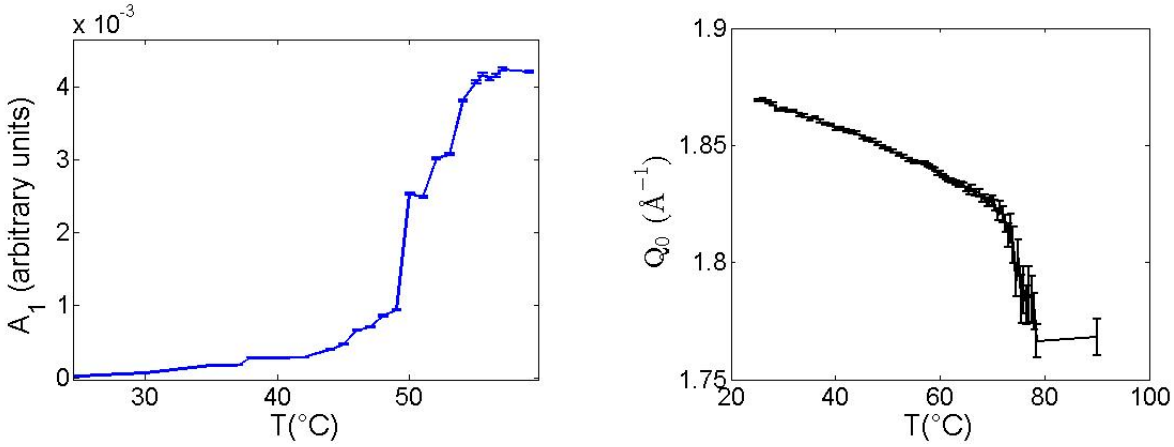
Table 4.7: Paramaters of the double-step function used to fit the data of the ethanol samples.

These parameters are qualitatively identical for all ethanol samples. The behavior of A_1 as a function of temperature for the ethanol samples is slightly different than for PEG samples (compare fig. 4.20a with fig. 4.15a). In the PEG case an almost linear increased from room temperature happened. For the ethanol samples A_1 stays roughly constant and increases steeply close to the melting temperature. Since the double-step function is a phenomenological approach to describe the diffuse contribution of submerged samples it is not straight forward to extract some physical meaning out of this differences.

On the contrary, $Q_0(T)$ for ethanol is qualitatively identical to what was observed in PEG samples (fig. 4.15b) with a steady decrease from room temperature due to thermal expansion and the relaxation of the base stacking along the molecule. The sudden drop at high temperature, whose significance is unclear, is also observed here.

Fig. 4.21 plots the A_0 parameter of the fit normalized by its value at room temperature for the ethanol samples. As in the case of PEG this parameter, which is proportional to the magnitude of the diffuse scattering at a given T, increases close to the melting transition for all the samples. For the ethanol samples this increase falls between 15% and 35%. This is consistent with the result for the PEG6K-17% sample which also was measured on D19. Therefore, the hypothesis that the changes in the magnitude of the increase of the diffuse scattering for different samples are related to the instrumental configuration (lack of collimation plus energy integration on D19) seems to hold.

A_0 for the ethanol samples does not show the non-monotonic variation of PEG6K-17% and PEG6K-20% (fig. 4.16). This non-monotonic behavior was ascribed to a coupling between the reversible changes in the diffuse scattering and the thermal oscillations of the



(a) (b)

Figure 4.20: Some parameters of the fit as a function of temperature for the Li-60% sample. a) A_1, b) Center of the Bragg peak (Q_0).

sample environment which were relatively big for these samples. The ethanol samples were measured with a thermal stability of ± 0.1 °C over all T, similar to the thermal stability of PEG8K-15% and PEG8K-20%. None of these samples showed non-monotonic behavior which reinforces the idea that thermal oscillations were causing the observed effect.

Fig. 4.22 shows the fitted integrated intensity (I_0) and width (γ) of the DNA Bragg peak; and the parameter A_1 of the double-step function as a function of the reduced temperature for the ethanol samples. Analogously to the PEG case, the I_0 of all the samples stayed roughly constant from room temperature to $T/T_m \approx 0.9$ and dropped steeply to values close to zero in the range $0.9 < T/T_m < 1.2$.

The I_0 of the Na-60%-0.1 sample shows a sharp peak just before plummeting ($T/T_m \approx 0.87$). At roughly the same temperature the γ of this sample has a dip. This sample is surely affected by some A-form contamination. It has been reported that heating up fiber samples with B-contamination can trigger a conformational transition which makes the B-form molecules go to A form [60]. If the opposite is happened in this sample (an A-to-B transition of relatively small parts) it will explain the origin of the observation. Base pairs in A-form molecules do not contribute to the 10^{th} layer peak (sec. 2.3.4), if the molecules

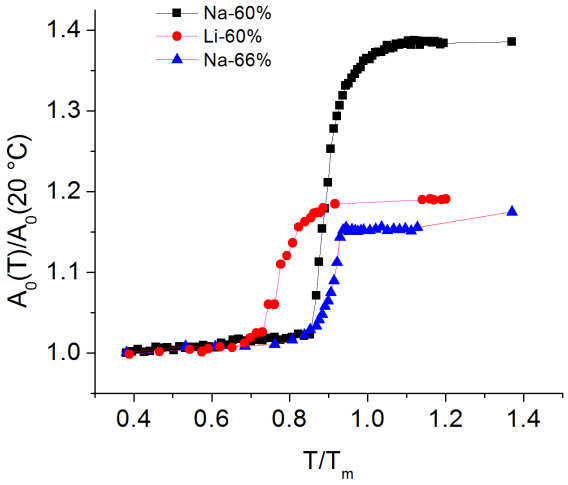


Figure 4.21: Parameter A_0 from eq. 4.5 as a function of the reduced temperature (calculated in Celsius) normalized by its value at 20 °C for each ethanol sample. A_0 is the average of the intensity at low Q and represents the magnitude of the diffuse contribution.

change to B-form at a given T their base pairs will contribute to the peak, increasing the I_0.

Note that these features can also be an artefact of the fit. The temperature at which the peak in I_0 and the dip in γ happened is also the T at which the background starts to change rapidly with temperature (A_1 starts to increase steeply with T). It may be that the fit, for this sample, cannot optimally determinate the ratio of the contributions from the Bragg peak and the diffuse scattering. The fact that similar features are not present in the Na-66%-0.01 sample, in which there is also A-contamination, reinforces the idea that the features are an artefact. However, to answer this question unambiguously, new fitting procedures must be tried and maybe more melting data in similar samples would be needed.

The matching of calorimetry and neutron data is satisfactory for the Na-66%-0.01 and Li-60% samples which reinforces the validity of the fits for these samples. The matching is not as good for the Na-60%-0.1 sample may be because of a relatively poor quality fit.

For the Na samples the width of the Bragg peak remained mostly constant until temperatures higher than T_m which is similar to what was observed on WOMBAT for the PEG samples. On the contrary, for the Li-60% sample a significant increase of the width starts

before the melting temperature ($T/T_m \approx 0.9$) is observed. This looks significantly different that the steady (linear like) increase of the width of the PEG sample measured in D19 (fig. 4.17a).

Discussion

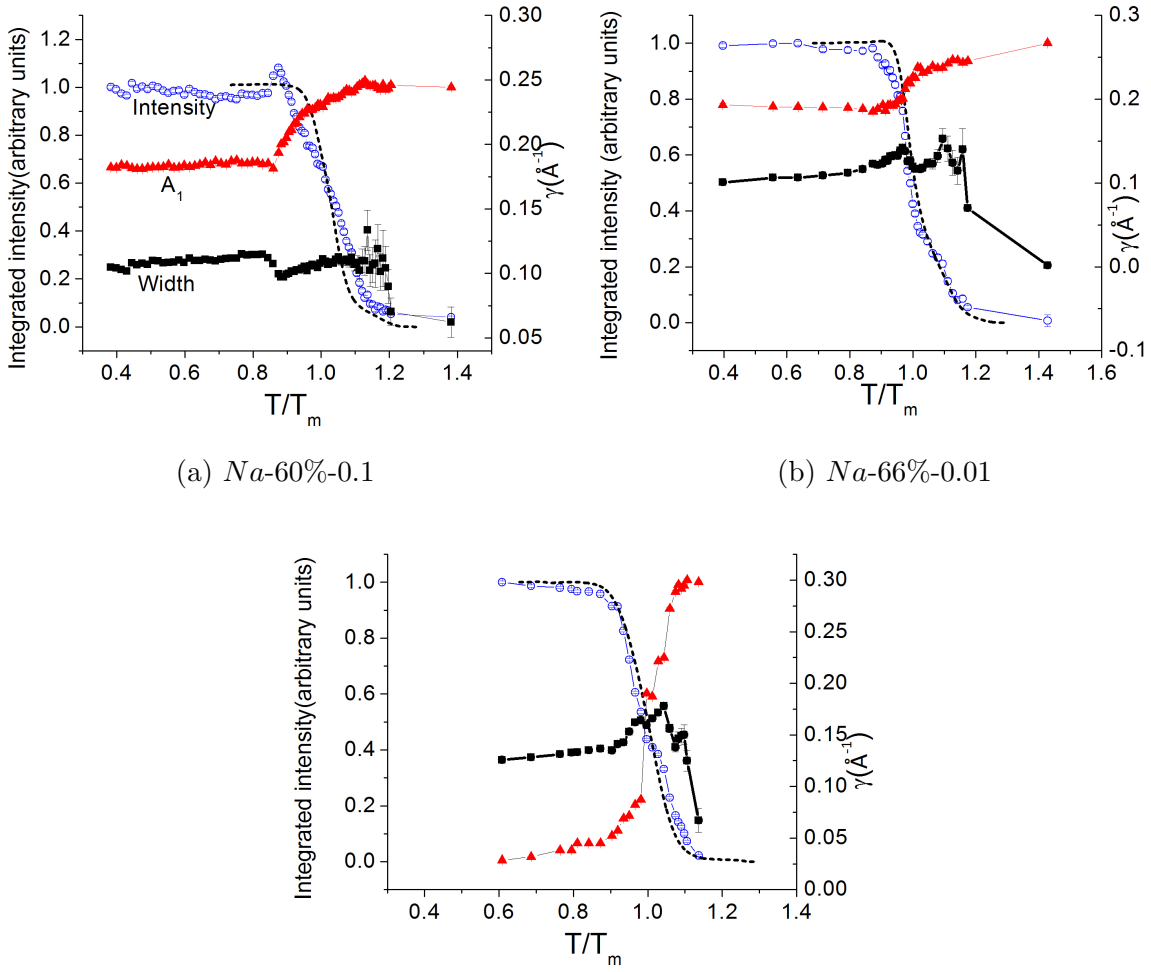
The experimental data presented in this section proves that the transition can be monitored equally well with neutron scattering in samples submerged in deuterated ethanol solutions with the advantage of the reduced incoherent scattering with respect to the samples with protonated PEG solutions.

The increase in the diffuse scattering (quantified by the parameter A_0 plotted in fig. 4.21) is qualitatively identical to the data of PEG6K-17% which was also measured in D19. This indicates that the large increase of the diffuse scattering in the PEG6K-17% sample as T raises is not a sample-specific behavior and reinforces the idea that the magnitude of this increase is linked to the instrumental configuration.

The double-step function seems to be in general less optimal to described the background of the ethanol samples. However, the fit of two samples (Na-66%-0.01 and Li-60%) is regarded as satisfactory and future attempts to reproduce the experimental data with the PBD model will be performed.

It is expected that the parameters necessary to model the data of these samples will differ significantly from the PEG. The effective potential of the binding between the bases (related with D_{AT} and D_{GC} parameters) is affected by the repulsion between the DNA strands which is reduced by the presence of ions around the DNA. This electrostatic effect is strongly affected by the dielectric constant of the solvent which changes with the amount of ethanol added to the solution.

Regarding the stacking of the bases, it is known to highly depend on the hydrophobic interactions [23] which must change in the presence of ethanol. For the Na-66%-0.01 sample the case is even more complex. A way to include the two peak melting in the relatively



(a) Na-60%-0.1

(b) Na-66%-0.01

(c) Li-60%

Figure 4.22: Parameteres of the fit as a function of temperature for the ethanol samples: integrated intensity (I_0, open blue circles) and width (γ, closed black squares) of the DNA Bragg peak placed at $Q_H \approx 1.87$ Å^{-1}; parameter A_1 of the double-step function (red triangles) which models the diffuse scattering. The dash black line represents the fraction of closed base pairs (f_{DS}) calculated by calorimetry with eq. 2.6.

simple PBD model must be found. This would be a opportunity to test the limits of the model.

Future experimental work will be focused on *Li*-DNA fibers since that will eliminate the A-contamination which may hinder the strength of the conclusions.

Overall, the temperature dependence of the ethanol samples is consistent with previous data but shows some specifities that, we believe, justify further experiments and data analysis in these kind of samples.

Chapter 5

Widom-601 investigated by SAS

5.1 Room temperature study

The bending and flexibility of DNA are of critical importance for many of its biological functions like gene regulation, protein-DNA interactions and DNA packing in viruses and eukaryote cells.

However the actual magnitude, nature and origin of DNA flexibility is still an open problem. Kahn has reviewed recently experimental works on bending and twisting of DNA which used DNA ring closure, atomic force microscope and optical tweezers methods [107]. Discrepancies in the results of the different techniques showed that the flexibility of DNA is far from being understood [107].

The magnitude of the flexibility of DNA is usually accounted for by the persistence length (l_p) which, within the Kratky-Porod model (section 3.4.2), is defined as the characteristic length in the correlation of the direction of the segments that describe the polymer. The l_p of DNA was initially thought to be ≈ 500 Å which would imply that DNA is hard to bend over shorter distances. However recent cyclization experiments [108], fluorescence resonance energy transfer measurements [109] and small angle X-ray scattering [110] performed on short

Figure 5.1: The 145 base-pair sequence investigated. The fragment in the rectangle is the strong positioning element, characteristic of this sequence. The TA fragments in red are the possible kink positions considered in the model.

chain DNA revealed that DNA is likely to bend spontaneously over much shorter distances than 500 Å.

The worm-like chain model is a continuous version of the Kratky-Porod model and has been widely used for describing long chain DNA. However, the previously mentioned studies suggest that the worm-like chain may not be applicable to short chain sequences. Also, these studies point to the possibility that the persistence length depends on the length of the DNA.

The close-packing of DNA inside the cell nucleus is a particular interesting event in which the flexibility of DNA is key to the biological function. The Widom-601 sequence, shown in fig. 5.1, has a strong histone positioning fragment which has great affinity to wind around a histone octamer to form a nucleosome which is a essential step of the packing of the genome in the cell nucleus.

Crick and Klug tried to address the mechanism behind the considerable folding the DNA undergoes to form the chromatin. They showed in a pioneering study that DNA can form sharp kinks without a major disturbance of its structure [111]. Indeed, Kinks have been

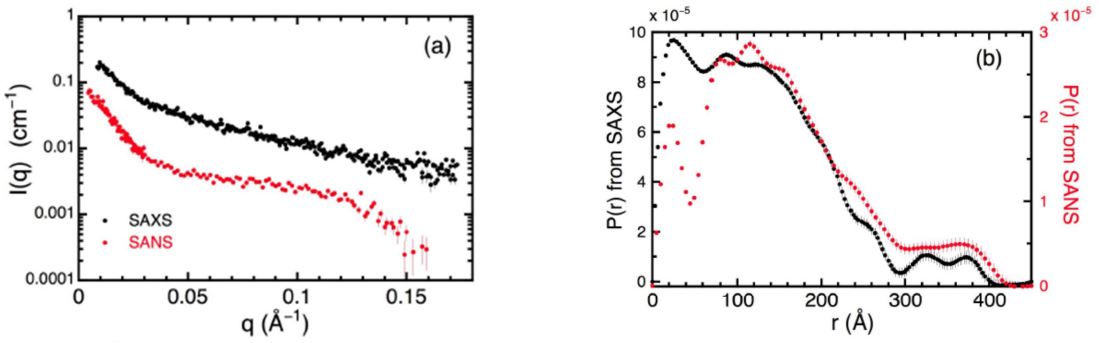
found in the structure of short DNA chains wrapped around a nucleosome [112] but Crick and Klug also hypothesized that kinks could occur in solution. This hypothesis was never definitively confirmed.

The Widom sequence has been previously studied under the effects of the strong interactions with histone proteins, [13] or attached to gold nanoparticles [110]. Investigation of the Widom sequence free in solution may provide the answer to the question of whether some intrinsic mechanical and/or geometrical properties of DNA are responsible for the strong positioning effect. The fundamental dilemma is whether the wrapping of DNA in nucleosomes requires a very high elastic energy, as the worm-like chain model suggests, or whether there is a sequence-dependent feature of DNA, like a enhanced local flexibility at the positioning site, that could explain or favour the positioning effect.

Therefore the aim is to study DNA molecules containing the Widom sequence dissolved in solution. Small angle scattering can probe the shape of the molecules dispersed in the solution and, when used with an appropriate model, the geometry and flexibility of the molecules can be estimated.

^2H_2O-based solutions of Widom-601 DNA molecules were prepared as described in sec. 2.1.2. Neutron and X-rays small angle scattering (SANS and SAXS respectively) data of these samples were recorded and reduced as described in previously (sec. 2.5.2, 2.4.1, 2.4.2 and 2.5.3). The use of the two techniques gave independent and complementary data. SAXS gives a larger signal from a much smaller mass of DNA but X-rays are well-known to damage DNA during long exposures. X-ray damage can hinder the experiment but it can also be used as a validation of the analysis from undamaged-DNA data as it will be shown below. Neutrons are non-ionizing radiation and the damage they cause to the sample is negligible. However, since the available fluxes are many orders of magnitude smaller, a greater mass of DNA is needed to obtained a usable signal.

Figure 5.2a presents the SANS and SAXS curves ($I(q)$) at room temperature. Fig. 5.2b shows the corresponding pair distribution function ($P(r)$), calculated as described in sec. 2.5.2.



(a) (b)

Figure 5.2: a) SANS and SAXS data at room temperature; b) their corresponding $P(r)$.

The $P(r)$ found by both techniques is qualitatively similar but there are some discrepancies. These are due to the differences in scattering length for neutrons and X-rays which arise from the different interaction mechanism of both radiations with matter (sec. 2.3.2). X-rays are more sensitive to heavier elements since the scattering length for them is proportional to the number of electrons in the atom. For neutrons hydrogen (which has a negative scattering length) has the highest contrast with the deuterated solvent and therefore the neutron data are more sensitive to the distribution of hydrogen in the sample.

The two data sets show peaks around 20 Å. For the X-ray data the peak is related to the phosphates in the outside of the double helix which are 20 Å apart. The DNA has many protons distributed along its structure, thus for short distances neutron scattering of DNA is well approximated by the scattering by a bulk cylinder with a diameter of 20 Å. Both $P(r)$ curves show a rather flat region between 100 and 140 Å. For $r > 140$ Å the $P(r)$ decrease steeply until they reach a plateau from 300 to 390 Å before going to 0 for $r > 400$ Å. The oscillations in the $300 - 390$ Å range in the SAXS data may be an artefact due to truncation in the measured Q range which entails missing Fourier components in the $P(r)$ calculation. However, the plateau is probably a real feature since it appears in both sets of data.

The calculation of $P(r)$ using the structure factor is a well-established procedure, which is based on a precisely defined mathematical connection between the functions (sec. 2.5.2). However, deriving the shape of the scattering molecules from $P(r)$ is not an unambiguous

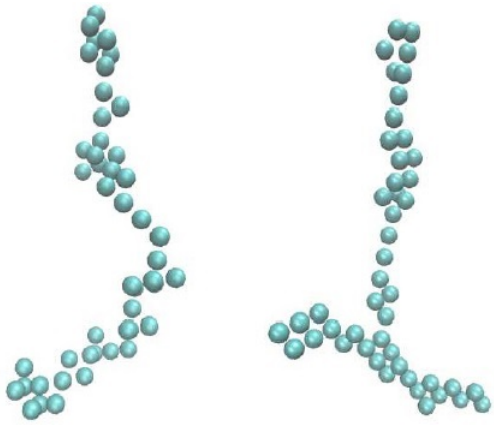


Figure 5.3: Representative examples of the output of the DAMMIF program. The dummy atoms show the estimation of the most likely shapes of the DNA in solution.

process. The DAMMIF program [113], included in the ATSAS package, was used as a first step. DAMMIF starts from the $P(r)$ function and uses simulated annealing to find the shape of a set of dummy objects which reproduce the $P(r)$. DNA can have a great variety of conformations in solution and multiple runs lead to different final shapes. DAMMIF converges to either strongly curved structures or branched solutions when the $P(r)$ shown in fig. 5.2b is studied. Fig. 5.3 shows representative examples of both cases. The in situ UV-Vis measurements collected in the SANS sample (sec. 2.4.1) ensures that the molecules were not partially denatured at room temperature. Therefore branching from individual molecules is not expected. The DAMMIF solution could still be explained if the sample contains a mixture of weakly curved and sharply bent molecules. In this case, the DAMMIF results would be due to a superposition of both conformations.

The ab-initio shape reconstruction that DAMMIF used is not conclusive because it does not include a priori knowledge of the molecular properties. In order to extract further information from the scattering data a theoretical model was developed.

An extension of the Kratky-Porod model ([96] and sec. 3.4.2) was used to analyze the $P(r)$ curves. The model describes the conformations that the backbone of the DNA can adopt but not the internal structure of the molecule. A schematic of the model used can be

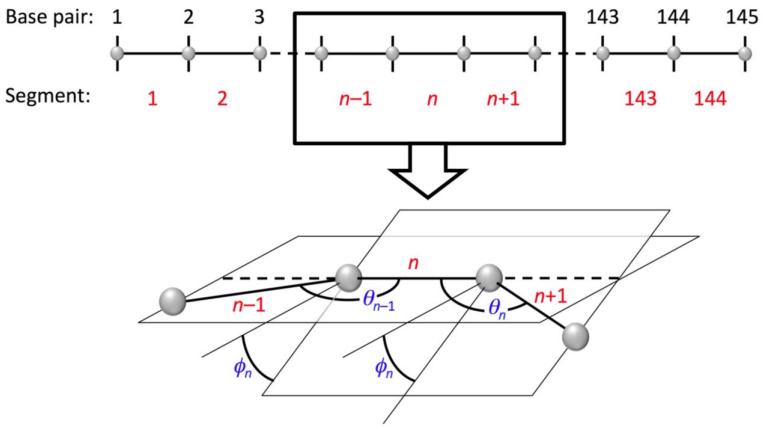


Figure 5.4: Schematic of the extended Kratky-Porod model used. The top part shows how the base pairs and segments were numbered and the bottom shows the definition for the bond and torsional rotation angles.

found in fig. 5.4. The model describes a single DNA molecule as consisting of $N+1$ objects representing base pairs. They are separated by N segments of length $a = 3.34$ Å which is the average base pair distance in B form DNA. The bond angles between the segments connecting the pairs may vary at each base pair. The local bending angle at site n, between segments $n-1$ and n, is θ_n. Torsional rotation is accounted for by means of a second set of angles, Φ_n, giving the dihedral angle between the plane containing segments $n-1$ and n and the plane containing segments n and $n+1$.

The model allows to compute a Hamiltonian which accounts for the thermal fluctuations by means of the bending and torsional energy. The Hamiltonian is given by:

$$H = \sum_{n=1}^{N-1} K_n[1 - \cos(\theta_n - \theta_{0n})] + \sum_{n=2}^{N-1} C_n[1 - \cos(\Phi_n - \Phi_{0n})] \tag{5.1}$$

where θ_{0n} and Φ_{0n} are local equilibrium values for the bending and torsional angles which can be different from zero, K_n and C_n are constants related with the bending and torsional energy. Dimensionless variables were used in the calculations with the model. Distances were measured in units of a, and temperature was expressed in energy units relative to k_BT

at room temperature. Thus $T = 1$ gives the properties of DNA at room temperature.

For any given conformation of the model a theoretical $P(r)$ can be determined by putting a unit scattering center at each base pair position. This is a good approximation for the SAXS data because the X-ray scattering from a base pair is dominated by the phosphorous atoms whose contribution (for $r > 40$ Å) can be effectively mapped onto the center of the base pair. Considering the base pair as a unit scattering center is a worse approximation for the neutron data because neutrons are more sensitive to the distribution of protons which is much more homogeneous throughout the molecule with respect to the phosphorous atoms. Therefore the theoretical $P(r)$ were compared to the SAXS data which, in addition, have better signal to noise ratio than SANS.

The data were first modeled with the classical Kratky-Porod model which has C_n=θ_{0n}=0 and assumes $K_n = K$ =constant for every segment. In this situation the persistence length is $l_p = Ka$. A persistence length of 500 Å corresponds to $K = 150$. Fig. 5.5 shows a comparison between the experimental data from SAXS at room temperature and the result of the calculation of $P(r)$ with these parameters (black line). This calculation has a very poor resemblance to the experimental data. The model gives a far better agreement if l_p is decreased to 60 Å. However, such a short persistence length seems fairly unrealistic. If the molecules were single stranded this persistence length could be adequate but the in situ UV-Vis absorption measurements performed confirmed that the sample was in helical form at room temperature. This suggests that, at least for short molecules (145 bp molecules), DNA cannot be modeled as a homogeneous Kratky-Porod polymer which agrees with the conclusions of the cyclization experiments [108].

It is not possible to perform a refinement of the model with all the parameters free because the parameter space is too large. A two-step process was adapted to circumvent this difficulty. First, a significant sample of geometric conformations able to reproduce the experimental $P(r)$ was found. Then, parameters able to generate these geometrical conformations when the model was thermalized were explored.

Monte-Carlo simulations were use to generate a large number of random conformations

and those which best matched the experimental $P(r)$ were selected. Since the generated conformations were totally random, some of them would be extremely unlikely. Therefore, an energetic criteria was used in order to avoid including in the study conformations that give the correct scattering pattern but are physically unrealistic . Some values must be selected for the model for this conformational search. They were: $K_n = K = 150$ (equivalent to assuming $l_p = 500$Å), $C_n = C = 2$, $\theta_{0n} = 0$ and $\Phi_{0n} = 0$. Every configuration generated was accepted with the probability $exp(-H/k_B T_0)$. T_0 was set to 3 so the conformational space could be explored widely.

A total of 12×10^7 conformations were generated. The $P(r)$ of the conformations accepted by the energy criteria were calculated and compared to the experimental SAXS data at room temperature by computing the standard deviation between the two functions. The 10^3 conformations with the smallest deviations were selected. The average $P(r)$ for these conformations is plotted in fig. 5.5 in red and labeled as Monte-Carlo.

The experimental and modeled curves match satisfactorily in the region 80 Å$< r <$ 280 Å. The calculation does not reproduce the peak around 20 Å which was expected since the model does not take into account the diameter of the DNA. However it shows a peak at ≈ 3 Å. This is related with the strong correlation between adjacent base pairs with are separated by $a = 3.4$ Å. The model also does not reproduce the plateau placed between 300 Å and 400 Å which shows that the model is not perfect. However, since the agreement is reasonable over a large range it may be assumed that the modeled $P(r)$ is a good representation of the state of the thermalized DNA molecules.

The 10^3 conformations were examined closely to find trends in the preferred shapes of the DNA molecules. A histogram of θ_n for the conformations can be found in fig. 5.6a, it shows that most of the angles for the different conformations are small (between 0 and 20°) but there is a significant contribution of very large values of θ_n around the center of the molecule. It is important to note that there is a gap between the large values of θ_n near the center and the small values at the same sites. If the large values were associated with a high local flexibility in the sites near the center, the distribution of values of θ would

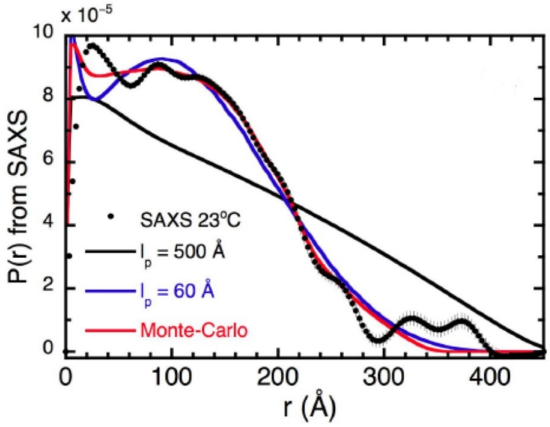
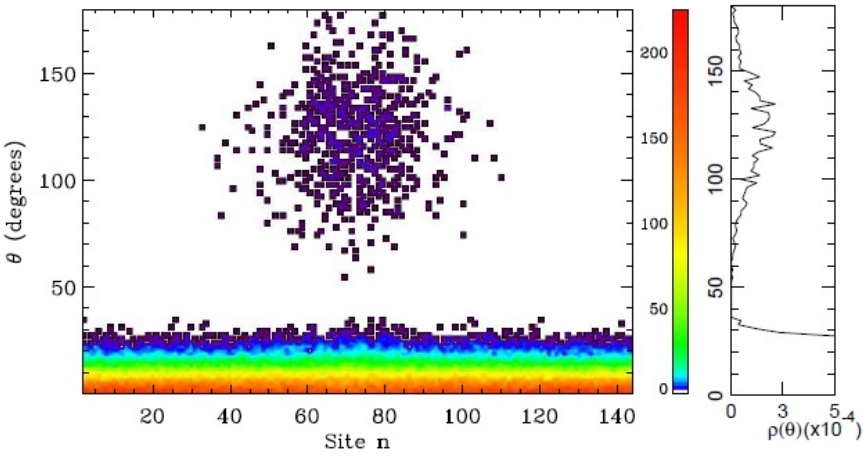


Figure 5.5: $P(r)$ calculated using the SAXS data (black dots), $P(r)$ calculated with the model of an homogeneous polymer with $l_p = 500$Å (black line) and 60Å (blue line) and average $P(r)$ from the 10^3 conformations with the best match to the experimental data from the 12×10^7 conformations generated with MonteCarlo.

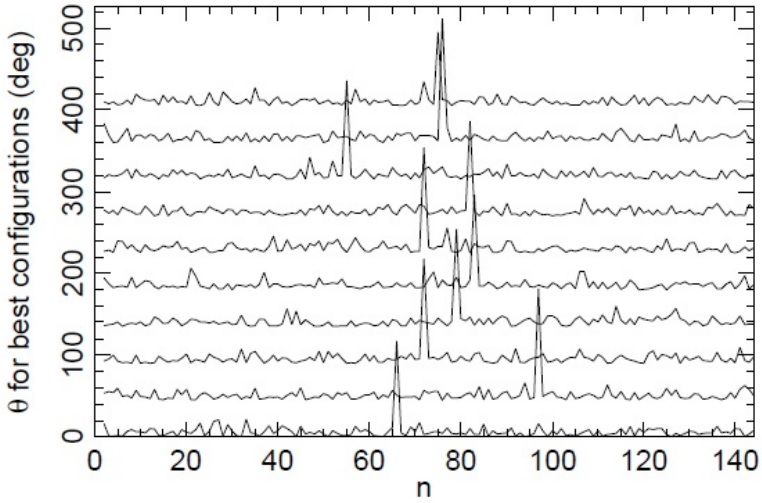
not show the peak in the right part of fig. 5.6a but would fall continuously from $\theta = 0$. The double peak structure suggest that the model has metastable states with large angles, which in turn suggest the existence of permanent bend for some conformations. Fig. 5.6b depicts the values of θ_n for the 10 conformations that provide the best matching with the experimental $P(r)$. It shows that, for these conformations, only one site has a very large θ_n while the other values are small.

The next step of the study was to select parameters for the model that included the main characteristics of the conformations which were able to accurately reproduce the scattering data. This procedure may confirm and clarify properties of the DNA molecules that were statistically tested during the conformational search.

The model used for the conformational search did not include any equilibrium bending or torsion angles for the majority of sites ($\theta_{0n} = \phi_{0n} = 0$). All sites were assumed to be equivalent in order to avoid any bias or a-priori assumption. However, the results of the conformational search suggested the DNA exibits a very sharp bend, of the order of 90° in the central region (Fig. 5.6a). This is consistent with the DNA structure and bonding if the



(a)



(b)

Figure 5.6: a) Histogram of the bending angle against n for the 10^3 selected conformations which provided the best matching to the SAXS data at room T with a color scale. The right part of the figure shows the fraction ρ_b of θ angles, integrated over the whole model, which belongs to a given range of θ. b) Values of θ_n for the 10 conformations which provided the best match to SAXS data at room T. Successive plots are moved up by 50 for the sake of clarity.

molecule has a kink as Crick and Klug described [111]. Within the model used here, a kink can be define by setting $\theta_{0n} = \theta_K$ at a specific site $n = n_K$.

The position of the kink, according to the conformational search, can be assumed to be close to the center of the molecule, specifically between sites 55 and 90 (fig. 5.6b). The most probable position can be further specified because the sequence of the DNA is known (fig. 5.1). Crick and Klug highlighted that kinks appear mainly between A-T pairs [111]. Structural investigations of the nucleosome and molecular modeling confirmed this hypothesis [114, 115]. In the Widom-601 studied here the TA steps closest to the center are located at $n = 47, 57, 67$ and 88, being $n = 88$ part of the strong positioning element TTTAA characteristic of this sequence.

Models based on the presence of kinks were used to try to reproduce the data. Some parameters remain unchanged with respect to the conformational search: $K_n = 150$, $C_n = 2$, and $\phi_{0n} = 0$. $\theta_{0n} = 0$ except at the kink site $n = n_k$, where the equilibrium bending angle was varied between models in the range $70° < \theta_K < 110°$. The probability distribution of this model for a given θ_K and n_K, $P_{theo}(r)$, was taken as the average of 5000 conformations calculated with Monte-Carlo when the model was thermalized at room T ($T = 1$). The branched results of the DAMMIF program suggested a possible mixture of kinked and non-kinked DNA molecules. This case was taking into account by computing:

$$P'_{theo}(r) = xP_{theo}(r) + (1 - x)P_{theo,0}(r) \tag{5.2}$$

where x is the fraction of kinked molecules and $P_{theo,0}(r)$ is the probability distribution of the distance in non-kinked DNA molecules, i.e. molecules with $\theta_{0n} = 0$ at all sites, at room T. $P'_{theo}(r)$ and $P_{theo}(r)$ were compared to the experimental probability distribution, $P_{exp}(r)$, of SAXS data by calculating:

$$\chi^2 = \frac{1}{N_r - 1} \sum_{1}^{N_r} [\frac{P_{theo}(r_j) - P_{exp}(r_j)}{\sigma_j}]^2 \tag{5.3}$$

n_K	$\theta_K(^\circ)$	χ^2 kink	χ^2 mix	x
88	95	2.31	1.09	0.80
88	110	9.17	1.62	0.62
88	80	1.26	1.26	1.00
88	70	3.00	3.00	1.00
67	**95**	**3.32**	**0.75**	**0.73**
67	110	9.17	1.62	0.62
67	80	0.88	0.87	0.99
57	95	2.14	1.24	0.82
47	95	2.91	2.86	0.84

Table 5.1: Results of the comparison between theoretical and experimental $P(r)$ using the results of 5000 Monte-Carlo simulations at room T for each couple θ_K-n_K. The table lists the distance χ^2 between the experimental and theoretical probability distributions (eq. 5.3) for kinked DNA and for a mixture of kinked and non-kinked DNA with x being the fraction of kinked DNA. The line in bold provides the best agreement with the experimental data and is plotted in fig. 5.7. For non-kinked NDA the χ^2-distance between the theoretical and experimental probability distributions is $\chi^2 = 22.86$.

where r_j are the points where $P_{exp}(r)$ was computed, σ_j is the error of $P_{exp}(r)$ given by GNOM at the point j^{th} and N_r the number of r_j points. This calculation is restricted to $r_j > 50$ Å. The polymer model does not take into account the diameter of the DNA molecule, therefore it cannot be expected to describe accurately the DNA molecules at small distances.

The results for models with different kink positions, kink angles and fraction of kinked molecules are shown in table 5.1. The best fit was found for a mixture of kinked with non-kinked DNA with $n_k = 67$, $\theta_k = 95\ ^\circ$ and $x = 0.73$, the parameters of this model are highlighted in bold in table 5.1. Fig. 5.7 shows the $P(r)$ calculated with this model in red. The results show that assuming the presence of a kink in the molecules allows a good fit of the experimental probability distribution over a broad range (50-300 Å). The polymer model cannot reproduce all the fine structure of the experimental data which is not surprising given the simplicity of the model. However, over most of the range in which the model applies

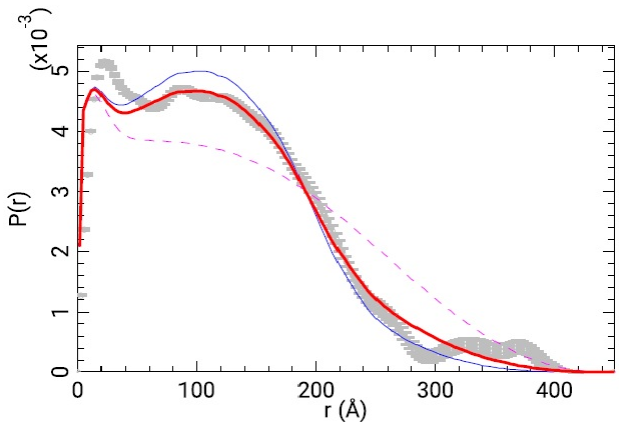
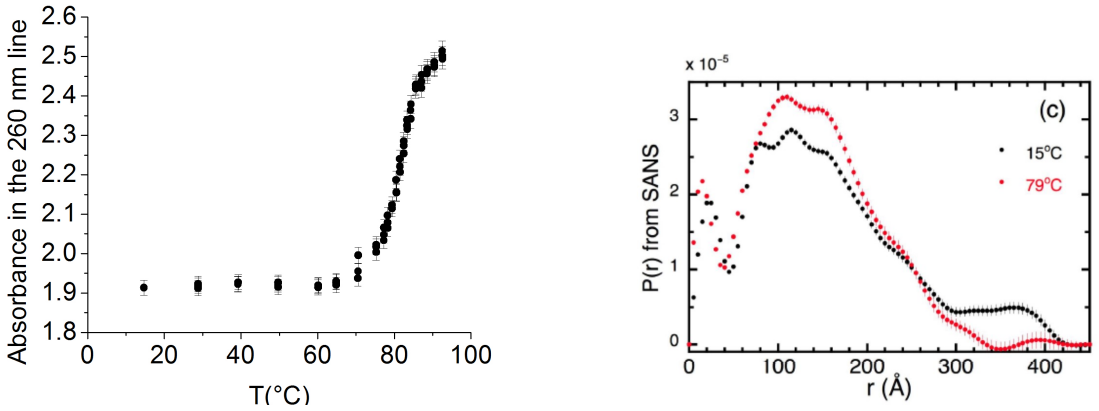


Figure 5.7: The grey points are the $P(r)$ from SAXS data, error bars are given by the GNOM calculation; the red line is the theoretical $P(r)$ obtained with a mixture of kinked ($\theta_K = 95°$, $n_K = 67$) and non-kinked DNA molecules with $x = 0.73$; the blue and magenta-dashed lines are the calculated $P(r)$ for the kinked DNA and the non-kinked DNA, respectively.

($r > 50$Å) the theoretical results fit the data within the experimental errors. The diameter of DNA should be included in the model to achieve accurate results at $r < 50$ Å. Regarding the discrepancy between experiment and theory for $r > 300$Å, it has been observed that introducing some extra flexibility near the ends (10 to 20 base pairs in each end) by reducing K_n improves the agreement in this range. This increased flexibility may be accounting for a partial unwinding of the helix in the free ends.

The success of the model which mixes the $P(r)$ of kinked and non-kinked molecules does not mean that there are two different populations of molecules in solution. It is far more likely that thermal fluctuations were flipping the molecules between a kinked and non-kinked state continuously. The gap shown in fig. 5.6a shows that the intermediate states during the flip go undetected which means that the lifetime of each state must be much longer than the duration of the flip.

The model which showed the best agreement with the experiment has $n_K = 67$ but the model with the kink at $n = 88$, within the strong positioning element, gives a result almost equally good. The analysis presented here cannot significantly discriminate between the two optimal kink sites. In addition, the site $n = 57$ is almost symmetric of $n = 88$ with



(a) (b)

Figure 5.8: (a) In situ UV-Vis absorption as a function of temperature. (b) $P(r)$ from SANS at 25 and 79 °C.

respect to the center of the molecule. For the SAXS/SANS experiment and for the model the two ends of the molecule are indistinguishable, therefore the 57 and 88 sites are almost equivalent. This is consistent with the fact that models with the kink at $n = 57$ give also good results. Wherever the kink was placed, the best agreement with the experiment was found with $\theta_K = 95$ ° in good agreement with the kink angle predicted by Crick and Klug [111].

The scattering from the DNA solution containing the Widom sequence was measured in function of temperature with in situ UV-Vis spectroscopy as described in previous chapters (sec. 2.4.1). Temperature-dependent data will be discussed extensively in the next section but it is interesting to highlight here some aspects of the data which reinforce the conclusions of the study with the model.

Fig. 5.8a displays the UV-Vis absorbance of the sample in the 260 nm line which is usually used as reference to follow the unstacking of the base pairs due to temperature [116]. Fig. 5.8b shows a comparison of the $P(r)$ of the SANS data at 25 and 79 °C. $P(r)$ clearly changes due to the increase in temperature, it increases in the range 80 to 200 Å and decreases to almost 0 for values greater than 300 Å. The UV-Vis data show that around 20% of the base pairs are opened at 79 °C. The changes observed with respect to room T are

related with the onset of the melting transition and an increase of the thermal fluctuations of the molecules.

For the SAXS data the measurements showed a dependence on the exposure time. This was not observed in the neutron data with similar exposure times. Fig. 5.9 shows three data sets, one at 23 °C and two at 70 °C. The three data sets were measured on 2 aliquots of the same sample. One aliquot was measured for one hour at room temperature to obtain the data set labeled as 23 °C. Then this aliquot was thermalized at 70 °C (with the beam shut down) and measured for another hour to obtain the data set labeled as "70 °C, short exposure". Therefore before recording the short exposure data the sample was exposed to X-rays for 1 hour. The second aliquot was measured for one hour at three different temperatures before being thermalized at 70 °C for measuring the data labeled as "70 °C, long exposure". Therefore, before recording the long exposure data the sample was exposed to X-rays for 3 hours.

The short exposure data (red line in 5.9) agree qualitatively with the high temperature SANS data of fig. 5.8b, showing changes with respect to room temperature happening in a similar range of Q with a similar magnitude. The long exposure data (blue line in fig. 5.9) are very different from both the short exposure SAXS data and the high temperature SANS data. It is believed that X-ray damage due to the relatively long exposure time is the cause. At X-ray fluxes as low as the one used in this study the main mechanism by which X-rays damage DNA is single strand breaking due to free radicals generated by absorption of the X-rays by the buffer [117]. Single strand breaking will significantly increase the flexibility of the DNA molecules which can explain the changes on the experimental $P(r)$ as a function of exposure time. This problem is not present in the neutron data since neutrons are not ionizing radiation.

A conformational search was repeated using $P(r)$ from the long exposure data (blue curve in fig. 5.9). The matching of the average $P(r)$ of the 10^3 selected conformations to the experimental data was again satisfactory and much better than trials with the classic Kratky-Porod model, even if torsional rigidity was included (fig. 5.10a). Fig. 5.10b shows

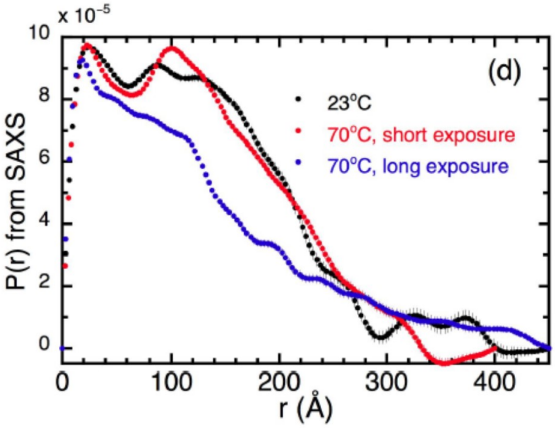
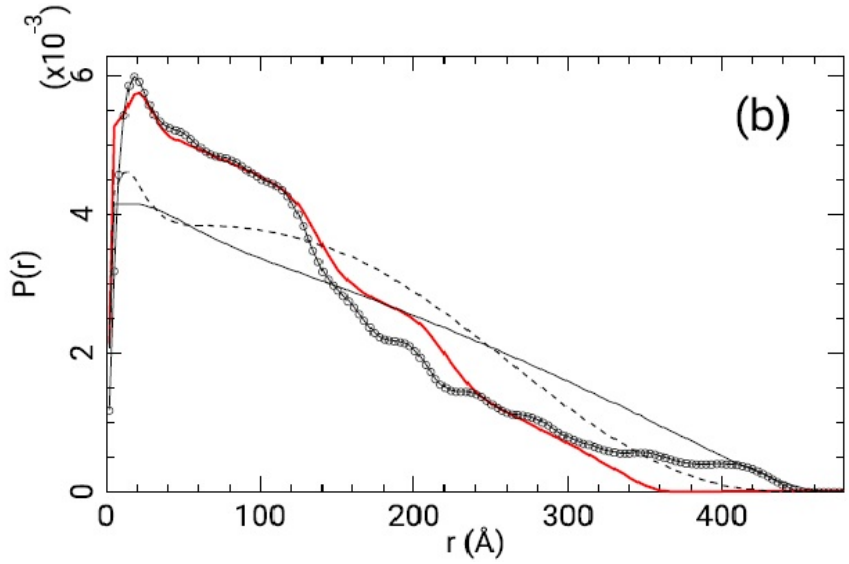


Figure 5.9: The $P(r)$ at 23 °C and 70 °C as a function of time of exposure to X-ray radiation. The short exposure data is the result of a measurement after the sample was exposed to X-rays for 1 h. The long exposure data is the result of a measurement on an identical sample after the sample was exposed for 4 h.

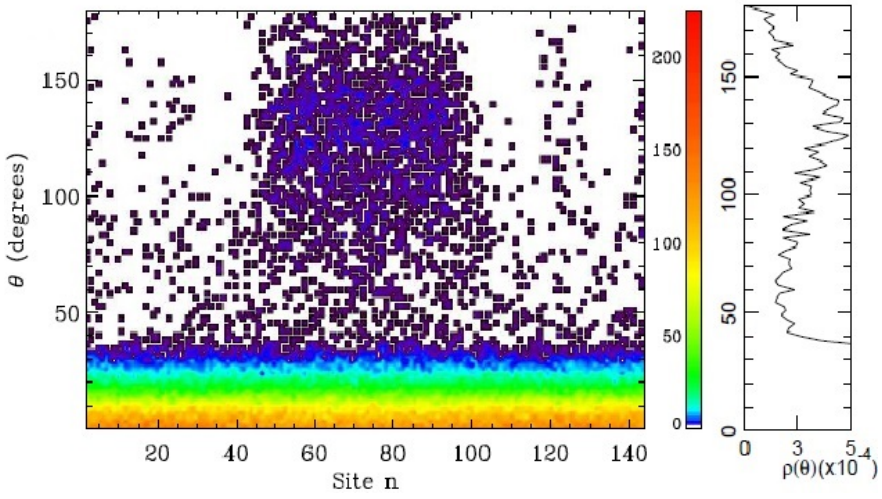
an histogram of the θ_n for the selected conformations.

In comparison to the room T histogram (fig. 5.6a) the histogram for these data shows many more sites in which θ_n can be large. Although there is still a gap between the large and small θ_n values, the distribution of θ_n is much more homogeneous than in the room T case. This can be explained by the single strand breaks caused by the X-ray damage which would increase the local flexibility dramatically and create a more continuous distribution of bending angles. The diagram on the right of fig. 5.10b still shows a maximum in the 90°-140°, therefore a kink seems to persist in these conditions.

In summary, the analysis of the scattering data suggested the existence of kinked Widom-601 DNA molecules in the solution at room temperature. The presence of a gap between the low and high bending angles near the center of the molecules, observed in the results of the conformational search, proves that an enhanced local bending flexibility of the DNA molecules is not enough to explain the scattering data. Complementary cryomicroscopy experiments [118], able to provide direct imaging, could perhaps confirm this result and validate further the kink hypothesis made by Crick and Klug.



(a)



(b)

Figure 5.10: a) Same as fig. 5.6a for the 10^3 conformations that provide the best matching with long-exposure SAXS data. b) Black circles are the SAXS $P(r)$ measured in a sample at 70 °C after a long exposures to X-rays; the red line is the average $P(r)$ for the 10^3 conformations which provided the best match to the experimental data; the full black line shows the $P(r)$ calculated with the polymer model with parameters: $K_n = 150$, $\theta_{0n} = 0$, $Cn = 0$; the dashed black line is the $P(r)$ for the same model if one takes into account the torsional rigidity ($Cn = 2$,$\phi_{0n} = 0$).

The results suggest that, besides energetic effects in the interaction between DNA and histone, mechanical effects may contribute to positioning. A kink in DNA would localize the histone octamer relative to the DNA and contribute to the binding. This investigation can also contribute to the open debate on the length-scale dependence of the DNA persistence length. Intrinsic curvature effects average out for long DNA molecules in experimental studies. However, that is not the case for short molecules and this can result in a reduction in the observed experimental persistence length of short DNA molecules.

5.2 Temperature-dependent study

It has long been known that the persistent length of DNA decreases as temperature rises [119, 120]. Theoretical efforts have attempted to link the temperature dependence of l_p with the thermally-induced base pair openings. Models which do not include the openings predict that l_p varies as $1/T$, which disagrees with experiments [120]. Theodorakopoulos and Peyrard have included fluctuation openings using a combination of the Kratky-Porod and the PBD models [15]. The DNA molecules were described with an inhomogeneous Kratky-Porod model with soft and hard joints corresponding to open and closed base pairs. Statistical information about the state of each pair was provided by the PBD model [9]. Such an approach can be used to study the sequence effects on local flexibility [121] which may have important biological consequences. Moreover, the link between fluctuational openings and local flexibility provides evidence that the dynamical character of DNA is very important for the description of the molecule, even at temperatures well below the melting transition.

Scattering data were used to try to experimentally test this theory. The samples used in sec. 5.1 were heated and their temperature-dependent small angle scattering was recorded, with in situ UV-Vis absorption spectroscopy for the SANS experiment. SAXS data presented here was recorded in samples which were exposed to X-ray for a maximum of one hour so the effect of the radiation damage can be neglected.

The l_p of the molecules can be estimated by fitting the small angle scattering curves

with an appropriate model. Theoretical predictions of l_p can be computed as in [15] with the advantage that the sequence of short artificial DNA is known and so the calculation of the statistical conformations of the molecules with the PBD model includes no arbitrary parameters.

5.2.1 SANS

Fig. 5.11a shows an example of the SANS curve of the DNA solution at room temperature. SANS data for the buffer solution, measured under identical conditions, are also shown. The signal from the buffer is featureless and the signal from the DNA appears mostly at low Q due to the significant size of the molecules (145 base pairs relates to a length of around 485 Å).

Fig. 5.11b shows representative examples of the SANS curves at low and high temperature. The most patent change with temperature is an increase of the magnitude of the intensity at large Q as temperature rises. Fig. 5.11c shows the mean intensity of the scattering for $Q > 0.17$ Å^{-1} for the DNA and over the whole Q range for the buffer plotted as a function of temperature. The buffer follows the same trend as the sample, so the increase in the scattering is not caused by the DNA molecules but by temperature-related changes in the solvent.

Fig. 5.11d show the representative curves normalized by the average intensity of the curve for $Q > 0.17$Å^{-1}. Once the difference in level is accounted for the curves show a quite similar shape even at the highest temperature reached (97°C) where the UV-Vis absorption indicates that around 90% of the base pairs are open (fig. 5.8a). With only 10% of the molecule in the helical form, the molecule is expected to be extremely flexible in comparison with the room temperature molecules.

Although not immediately obvious, the DNA signal does change with T . Fig. 5.12a shows the result of dividing the data at each temperature by the data at 15 °C . The divided data were fitted with straight lines. The gradients of the lines clearly change with T. Fig.

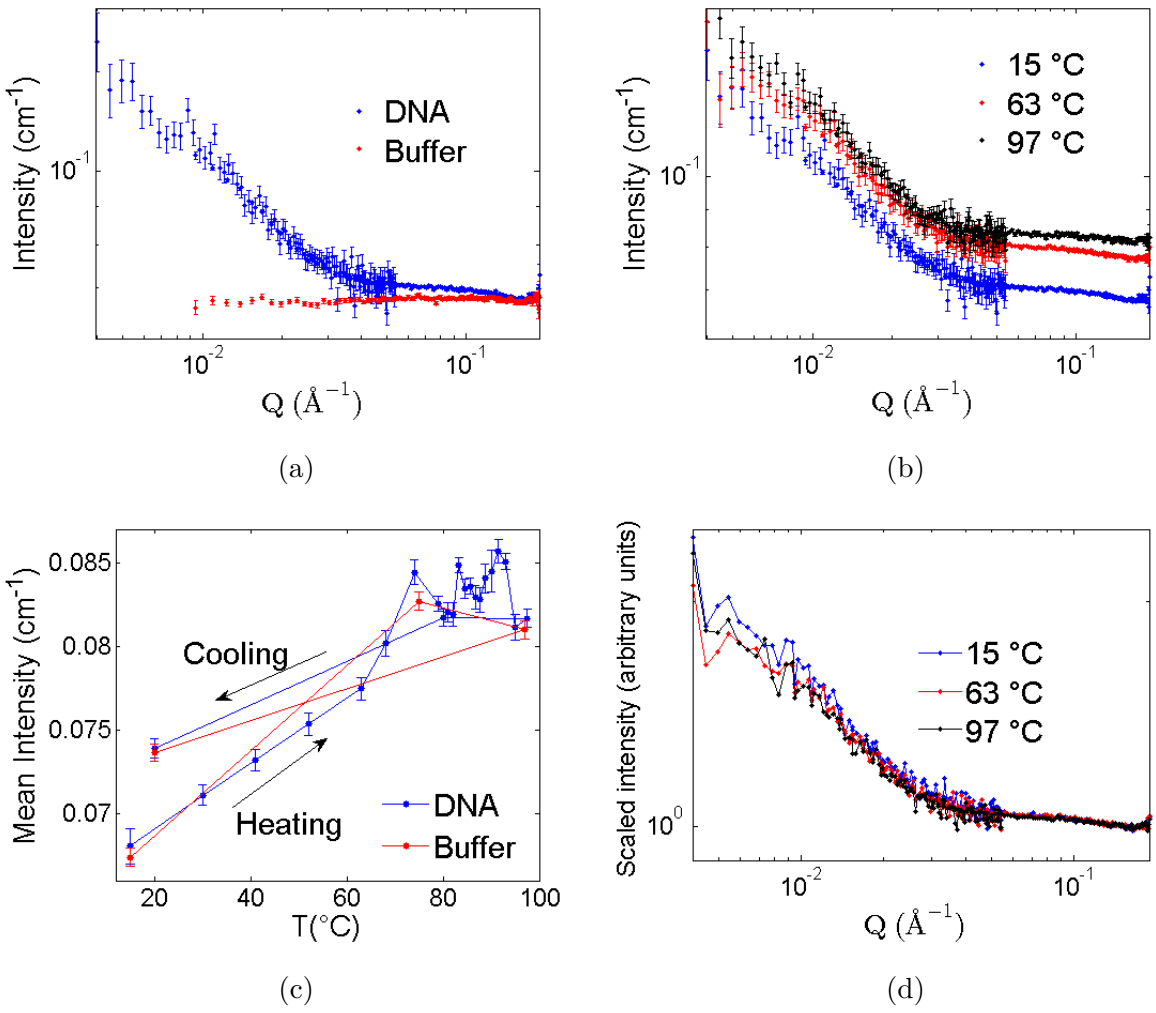


Figure 5.11: a) SANS curve of the DNA solution and of the buffer the DNA was dissolved in at room temperature in log-log scale. b) Representative SANS curves of the DNA solution at different temperatures in log-log scale. c) Value of the average intensity for $Q > 0.17$Å^{-1}for the DNA sample and over the whole Q range for the buffer as a function of temperature. d) Same curves of panel b) normalized by the average intensity of the curve for $Q > 0.17$Å^{-1}, error bars were removed for the sake of clarity.

5.12b shows the gradient of the linear fits as a function of temperature alongside the UV-Vis absorbance of the DNA.

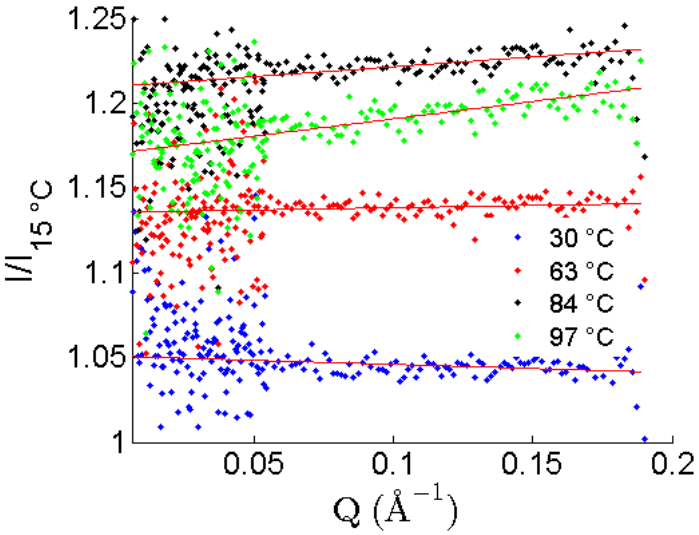
The gradient increases with increasing temperature. This shows that scattering at larger Qs increases relative to the signal at low Q as temperature raises, therefore smaller length scales are more dominant in the average structure of the molecules. This could be due to a bending over short spatial lengths which is consistent with an increase in local flexibility of the molecules and thus with a decrease of l_p. The fact that the U-Vis absorption seems to be correlated with the behavior of the gradient suggests that this effect is related with the pre-melting and melting of the molecules.

Fits to the small angle scattering data as a function of the temperature were attempted using the commercial software package Sasview [122]. The model used was called "flexible cylinder" and it describes the DNA molecule as a Kratky-Porod like chain (sec. 3.4.2) but the base pairs are represented as hard spheres (no overlap allowed) of radius R instead of dimensionless points. Thus the model accounts for self excluded volume effect when considering the conformations the molecule is allow to adopt and also for the effect of the radius of the DNA in the scattering pattern.

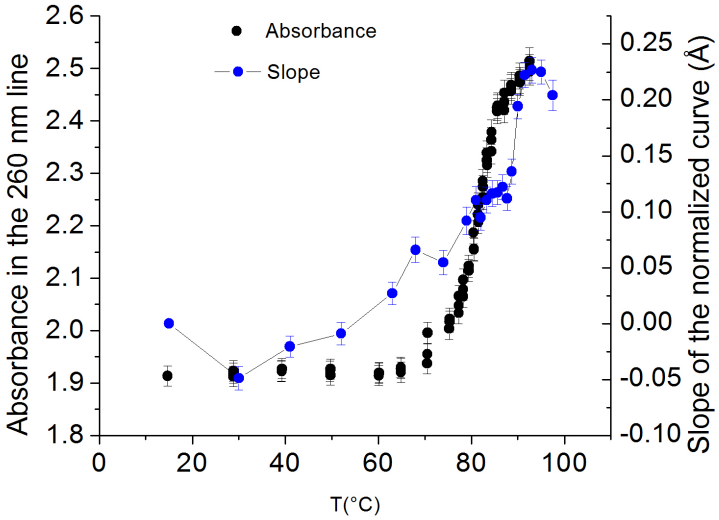
The derivation of the fitting function is based on Monte-Carlo simulations which used this model. Pedersen and Schurtenberger fitted the simulations and used the results along with numerical approximations to derive a phenomenological expression of a scattering function for semiflexible chains ($S_{WC}(Q, L, l_p)$) which accounted for the excluded volume effect [123]. If the L of the chain is significantly larger than the R of the hard spheres the scattering can be expressed as a multiplication of two terms using the decoupling approximation [124], thus the scattered intensity from a DNA solution is:

$$I_{WC}(Q, L, l_p, R) = D_s[(SLD_{DNA} - SLD_{sol})^2 S_{WC}(Q, L, l_p) P_{CS}(Q, R)] + I_b \tag{5.4}$$

where D_s is an scaling factor proportional to the DNA concentration, SLD_{DNA} is the scattering length density for the DNA molecule , SLD_{sol} is the scattering length density for

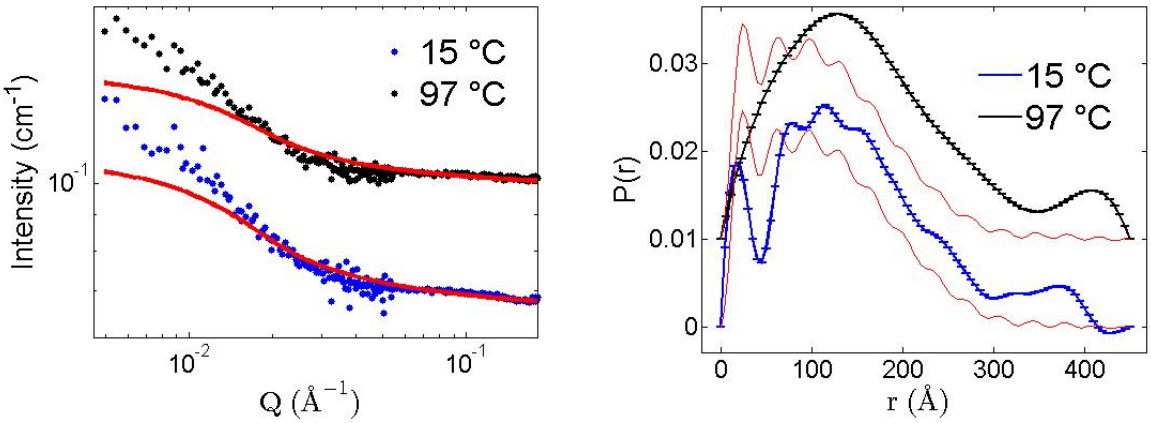


(a)



(b)

Figure 5.12: a) Examples of the result of normalizing the SANS curves of different temperatures to the data at 15 °C along with the linear fits (red lines). b) Slope of the linear fit to the normalized data and absorption of the sample as a function of temperature



(a) (b)

Figure 5.13: Examples of the fit to the SANS data at 15 °C and 97 °C. a) Intensity vs Q in log-log scale. b) Experimental $P(r)$ and $P(r)$ of the fits, they have been shifted vertically for the sake of clarity.

the solvent, I_b is a constant background, $S_{WC}(Q, L, l_p)$ is described in [123] (section Method 3 with excluded volume) and S_{CS} is the scattering function from the cross section of a rigid rod [125]:

$$S_{CS}(Q, R) = [\frac{2J_1(QR)}{QR}]^2 \tag{5.5}$$

where J_1 is a Bessel function of the first kind.

The instrumental smearing for SANS was accounted for by using dQ values for each Q which represented the standard deviation of Gaussian functions and were calculated as explained in [46]. The SLD for the solvent for neutrons ($SLD_{sol,N}$) was fixed to 5.754×10^{-6} Å^{-2}, the value of ^2H_2O [126]. R was fixed to 10 Å and L to 485 Åwhich is reasonable for DNA molecules of 145 base pairs in solution. D_s was fitted at room temperature, the value found was 58.23 ± 9.26. This value was fixed for the fit at other temperatures. Three parameters were fitted at every temperature l_p, $SLD_{DNA,X}$ (SLD of DNA for X-rays) and I_b.

Fig. 5.13a shows the fit to the SANS curves. The fit matches the data at high Q but is not good for low Q. Since l_p is a property of the overall molecule, low Q values are more important for an estimation of the persistence length. Fig. 5.13b shows the $P(r)$ of the experimental curves and the $P(r)$ obtained from the fitted curves given by Sasview. The matching of data and fit in this representation is quite poor. The value of l_p found by the fits do not vary (within error bars) with temperature: $l_p = 180 \pm 12$ Å. $SLD_{DNA,N}$ (SLD of DNA for neutrons) also remained constant with T within error bars; with $SLD_{DNA,N} = 5.76 \times 10^{-6} \pm 4.7 \times^{-11}$. This value of $SLD_{DNA,N}$ is just slightly bigger than the $SLD_{sol,N}$ which indicates a low contrast between DNA and solvent. This is consistent with the low signal of the sample detected during the experiment. I_b is related with a constant background and its behavior with T was almost identical to the mean intensity at high Q plotted in fig. 5.11c as expected.

However, a temperature-independent l_p is not consistent with previous works ([119], [120]) and looks unreasonable providing that at the highest temperature most of the molecule was melted as indicated by the UV-Vis absorption data (fig. 5.8a). Most probably the quality of the data or/and the accuracy of the fit are not good enough to detect a small change in persistence length caused by the increase in temperature.

5.2.2 SAXS

Fig. 5.14a shows that the SAXS data do not undergo the same increase in high-Q scattering with temperature observed with neutrons. Under close examination an increase at high Q can be observed in the 92 °C data with respect to the room temperature but the magnitude is much smaller for the SAXS data in comparison to the SANS data. In addition, a patent change in the shape of the curves as temperature increases is observable in these data.

Fig. 5.14b shows examples of a division of the SAXS data by the room temperature data analogously to fig. 5.12a. They were fitted with straight lines to compare to the neutron results. Fig. 5.14c shows that the temperature dependence of the gradient for SAXS data has similarities with the gradient of the neutron data and also correlates with the absorbance

measured in the SANS sample. Thus both X-rays and neutrons suggest qualitatively increase bending over small length scales as temperature rises which is consistent with an increase in flexibility.

The same fitting function and procedure presented for the SANS data was performed on the SAXS data as a function of temperature. For the SAXS data, the instrumental smearing was accounted for by assuming the resolution is defined by the collimation slits of the SAXS device ($0.4 \times 0.4\ mm^2$). L and R were fixed to 485 Å and 10 Å as in the SANS case. The SLD for the solvent for X-rays ($SLD_{sol,X}$) was fixed to the value of ^2H_2O, 8.508×10^{-6} Å^{-2} [126]. The scaling factor D_s was fitted at room temperature, the value found was 64.32 ± 12.14. The free parameters for the fit were l_p, $SLD_{DNA,X}$ (SLD of DNA for X-rays) and I_b.

Fig. 5.15a shows representative examples of the fit of the SAXS data at room temperature and the highest temperature studied using Sasview and the flexible cylinder model. The fit matches the data remarkably well. Fig. 5.15b shows the same fit in $P(r)$ representation. The matching is reasonable but the $P(r)$ calculated with the fitted curves do not follow exactly the experimental $P(r)$ possibly because the experimental data has a statistical variation which is lacking in the fitted curve.

Fig. 5.15c shows the fitted I_b and $SLD_{DNA,X}$. $SLD_{DNA,X}$ is constant with T within error bars and approximately 9.45×10^{-6} Å^{-2} which is slightly greater than the $SLD_{sol,X}$. This is consistent with the low signal detected during the experiment which suggest a small contrast between DNA and solvent. I_b has a value close to zero through out the whole studied temperature interval. It increases significantly at the highest temperature. This is consistent with the observed behavior of the data.

Fig. 5.15d shows the fitted l_p as a function of temperature. The value of l_p able to accurately fit the data at room temperature was 122.50 ± 26.51 Å. This value is more reasonable than the 60 Å needed with the classical Kratky-Porod model (fig. 5.5). The Kratky-Porod model does not account for the radius of the molecules and seemingly this has a significant effect in the flexibility detected by the model. However, 122.5 Å for l_p seems still unrealistic taking into account the structure of the double helix with a diameter of 20 Å. This value is

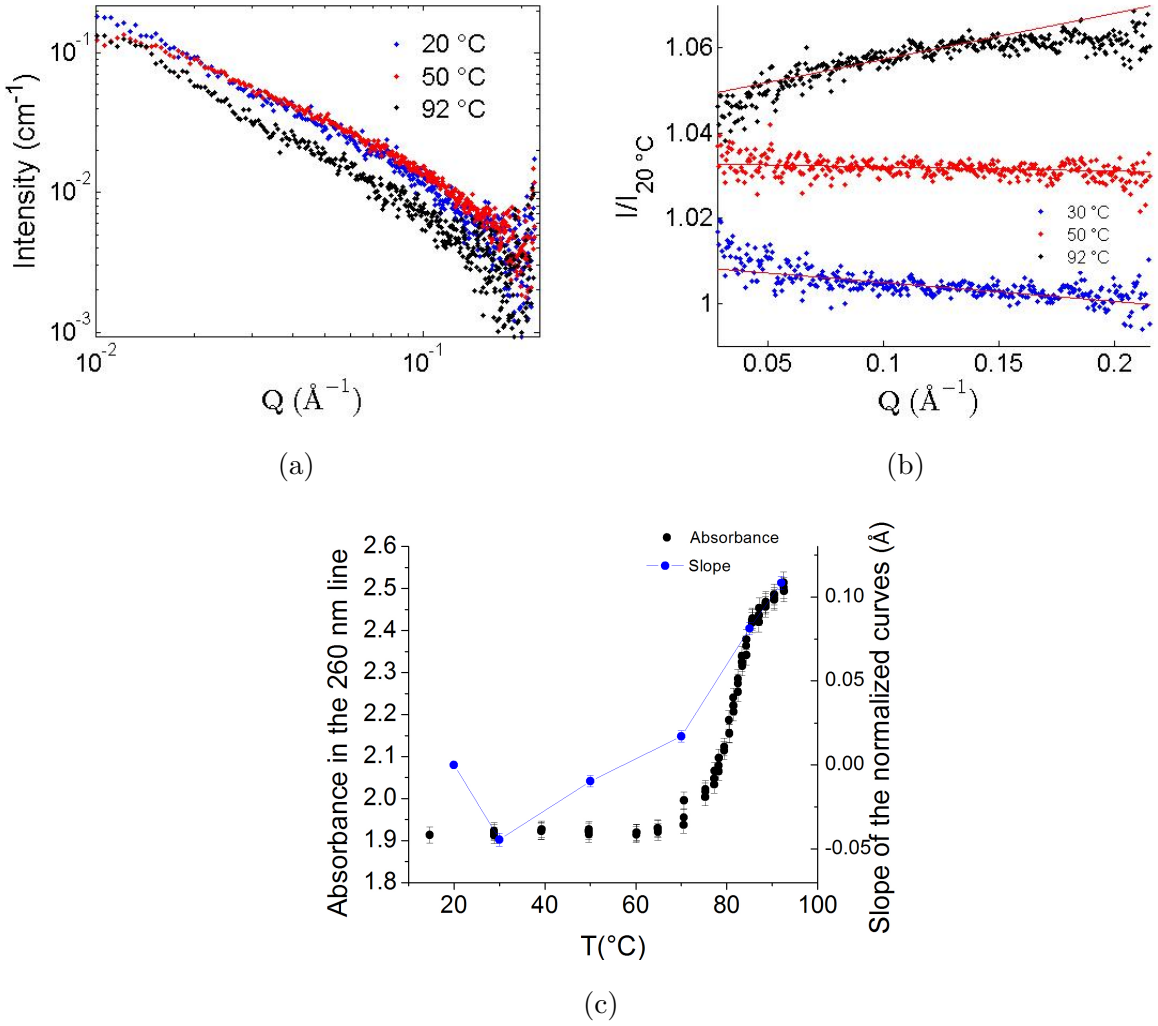


Figure 5.14: a) Representative SAXS curves of the DNA solution at different temperatures. Errorbars are not presented for the shake of clarity. b) Examples of the result of normalizing the SAXS curves of different temperatures to the data at 20 °C along with the linear fits (red lines). Curves have been shifted vertically for the sake of clarity. c) Slope of the linear fit to the normalized data and absorption of the sample as a function of temperature. There was not in situ absorbance measurements for the SAXS set up, the absorbance is the one of the SANS sample.

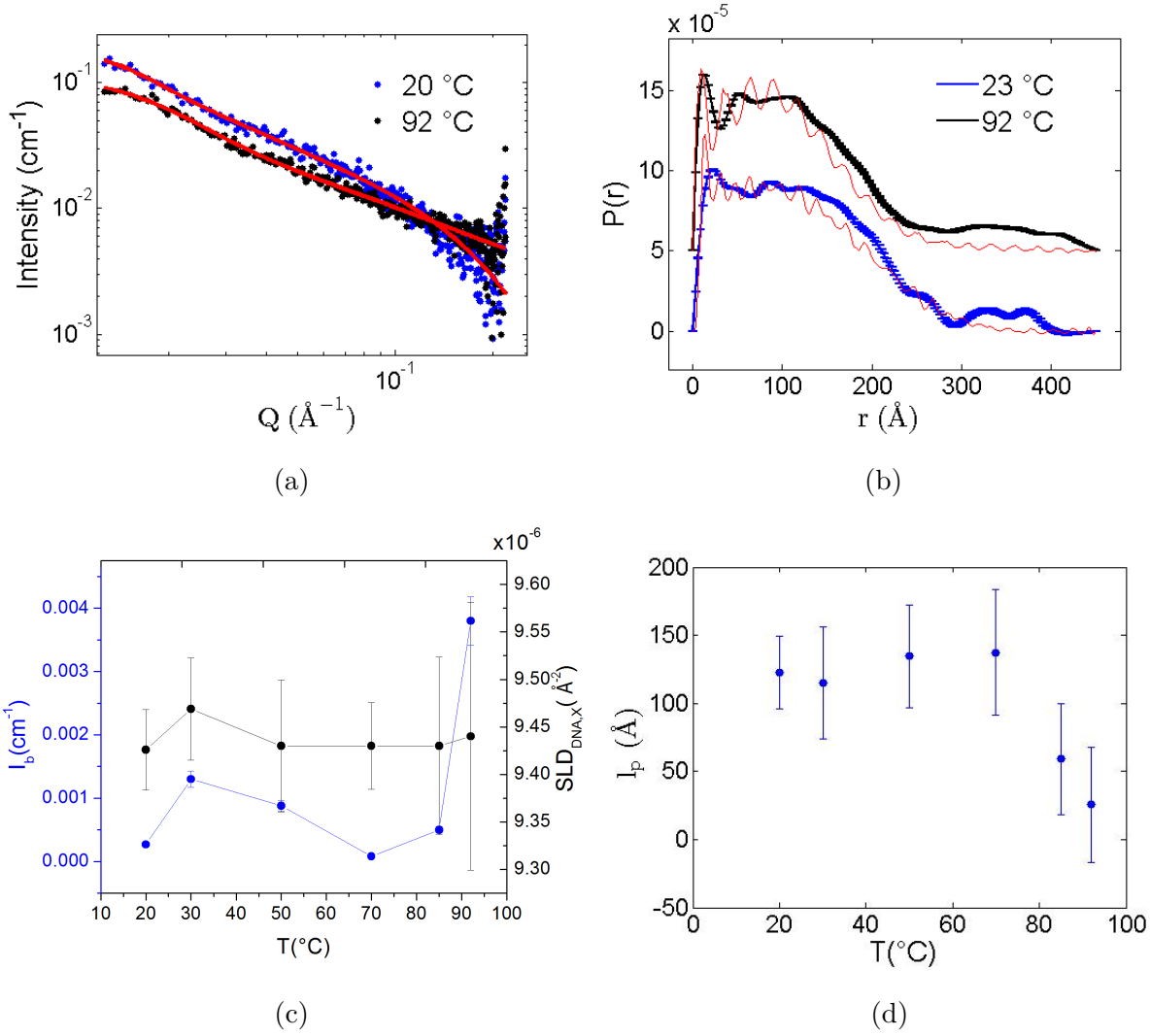


Figure 5.15: a) Examples of the fit to the SAXS data at 23 °C and 92 °C(Intensity vs Q). b) The same fits in $P(r)$ representation, the curves have been shifted vertically for clarity. c) Fitted I_b and SLD_{DNA} as a function of temperature. c) Fitted l_p as a function of temperature.

also three times smaller than measurements in single molecules [109] that were themselves significantly below the generally accepted value for DNA. Maybe this short persistence length is not a reflexion of a very high flexibility of this molecules but of the effect of the kink whose existence was strongly suggested by the room temperature study of previous sections.

Regarding the temperature dependence of l_p, the value of l_p found by the fit decreases from 70 °C on. Following the UV-Vis absorption (fig. 5.14c), 70 °C is roughly the onset of the melting so this suggest that the opening of the base pairs and the temperature dependence of l_p are correlated. This set of data defends qualitatively the idea of the base pair openings being a major factor in the temperature dependence of l_p but the uncertainties ascribed by the fit to l_p are almost as big as the change detected with temperature so a quantitative comparison with the theory as done in [15] would not be very meaningful.

5.2.3 Discussion

The definition of radius of gyration was introduced in previous sections, along with the persistence length, with regard to the classic Kratky-Porod model (sec. 3.4.2). The fact that the neutron data shows very subtle change with temperature and that the fit cannot estimate the values of l_p without large uncertainties for X-rays can be explained if the relation between persistence length and radius of gyration is examined closely. Fig. 5.16 shows the radius of gyration as a function of the persistence length. Both quantities are normalized by the length of the molecules. Assuming that the value out of the SAXS data for l_p is accurate ($l_p \approx 122.5$ Å) at room temperature, $l_p/L = 122.5/480 \approx 0.25$ and the expected R_g/L of the molecules studied is approximately 0.22 which is highlighted with a horizontal red line in the figure. In this range a doubling in l_p results in only a 10% change in R_g. Small angle scattering techniques are more sensitive to the overall conformation of the molecules in solution (directly related with R_g) than to changes in the internal structure of the molecules. This suggest that, for the DNA molecules studied, even if the persistence length is changing significantly as temperature rises the scattering data will respond only slightly. Therefore it will be extremely challenging to improve the accuracy of the fitting of l_p so that a quantitative

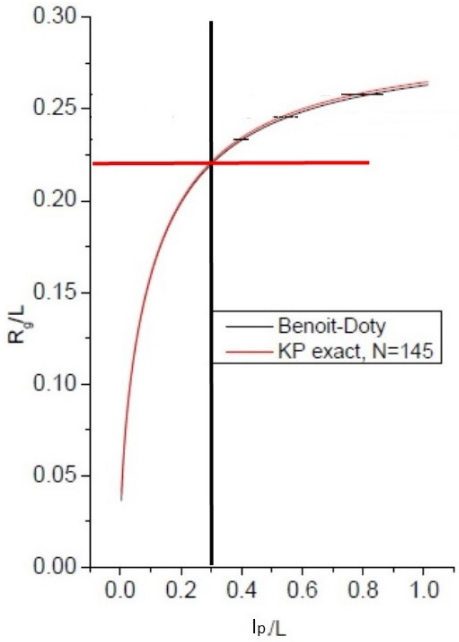
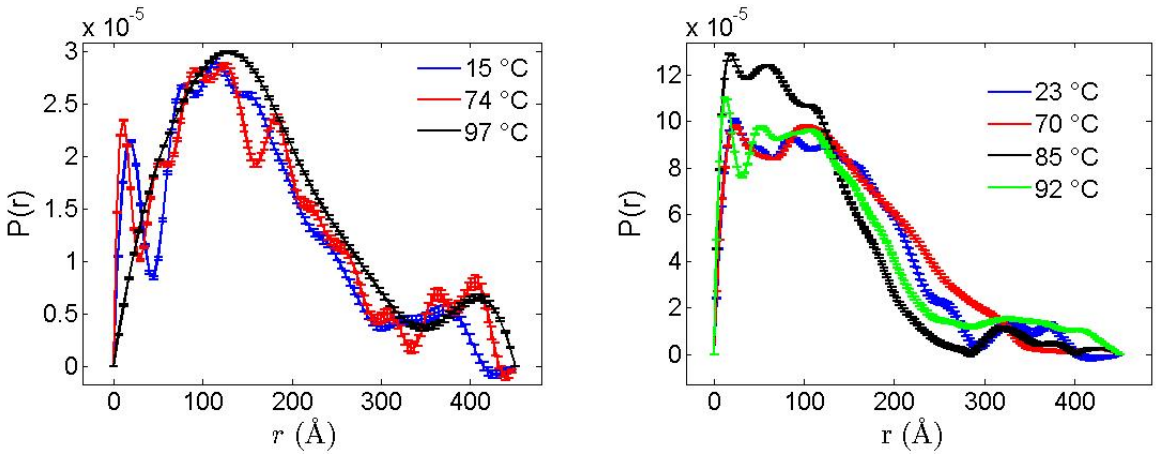


Figure 5.16: Relationship between R_g and l_p normalized by the length of the molecule calculated by the discrete Kratky-Porod model and the Benoit-Doty equation [127]. Figure provided by Nikos Theodorakopoulos.

analysis can be performed.

R_g and l_p are characteristics of the overall molecule and thus are strongly influenced by low Q data. The obstacle of the weak dependence of R_g on l_p for the DNA molecules under study can be avoided if a method that uses the whole Q range equally to access information is used. A possible avenue to extract further information about the T-dependent flexibility of DNA from the SANS and SAXS data is to repeat carefully the analysis of sec. 5.1 at each temperature in the data which did not suffer radiation damage due to the relative short exposure to X-rays. As described previously this method starts by generating random conformations which are energetically favorable, calculating the $P(r)$ of such conformations and comparing them to the experimental $P(r)$.

Fig. 5.17 shows representative examples of the calculations of $P(r)$ for SANS and SAXS data as a function of T. For neutrons the changes in $P(r)$ with T are subtle until relative high T. However, for the highest temperature the $P(r)$ changes abruptly, this may be enough



(a) (b)

Figure 5.17: a) Representative examples of the $P(r)$ of the SANS data as a function of temperature. Errors are given by the GNOM program. b) Same for the SAXS data.

to detect a significant difference during the conformational search. For the $P(r)$ calculated using SAXS the changes with T are dramatic and most probably the application of the model will give useful information.

In summary, the small angle scattering data recorded as a function of temperature suggest qualitatively that the dependence of the flexibility of short DNA with temperature is strongly influenced by the thermally induced openings of the base pairs. Extracting quantitative information by a fitting of the $S(q)$ to determine l_p has proven to be challenging. A better option seems to be to apply the extended Krakty-Porod polymer model described in sec. 5.1. This may be addressed in future works.

Chapter 6

Conclusions

This thesis tries to broaden the available knowledge about the DNA molecule, specifically its melting transition, by following two avenues: the study of highly oriented DNA fibers submerged in PEG or ethanol solutions; and the study of a short DNA sequence with biological significance (Widom-601 sequence) suspended in solution.

In chapter four, the room temperature study using neutrons on the samples submerged in PEG solutions showed that the long range order of the submerged fibers is decreased, possibly due to partial solvation and to the swelling of the fibers. However, the remained orientation is still sufficient to give relative strong Bragg peaks. The disposition of these Bragg peaks indicated that all the PEG samples were in pure semicrystalline B form. Thus, this kind of samples constitute a highly suitable model to study DNA-aggregates in a highly hydrated stated using scattering techniques in which long range order facilitates access to the fine structure of the molecules. The data recorded in the transversal orientation allowed the intermolecular distance to be estimated. For a given molecular weight PEG, this distance increased as the PEG concentration decreased. Since the other parameters of the solution remain mostly constant with PEG concentration, the submersion of the fibers in PEG solutions allows to study the properties of DNA as a function of the confinement or DNA-DNA interactions.

The calorimetry data of PEG samples showed that the melting temperature increases with the concentration of PEG1K, PEG6K and PEG8K. The relation T_m vs [PEG] is identical (within error bars) for PEG6K and PEG8K and differs appreciatively for PEG1K. There is a qualitative difference between the effect of PEG1K and the other larger PEGs in the width of the transition (proportional to the cooperativity).

The differences were explained as the PEG1K penetrating the DNA fibers and affecting the transition through a excluded volume effect, analogously to what would happen in a very concentrated solution of DNA and PEG. On the contrary PEG6K and PEG8K were believed to remain outside of the fibers. Their main effect in the parameters of the transition seems to be the osmotic pressure of the solution modifying the interaxial distance between DNA molecules, which in turn changes the DNA-DNA interactions. The data was compared to theoretical calculations of the intermolecular interactions as a function of interaxial distance. A high correlation between changes in the interaction forces, T_m and width of the transition was found (in the range of applicability of the model).

In the temperature-dependent study of the PEG samples using neutron scattering, the melting transition was successfully monitored by following the evolution of the Bragg peak. This can be asserted because the number of closed base pairs as a function of temperature detected by neutrons ($\propto I_0$) was consistent with calorimetry measurements. The evolution of the average size of the closed domains for all samples ($\propto \gamma$) proved that large domains of closed base pairs remain until temperatures very close to or even higher than the melting temperature. This result is consistent with previous studies from the group on the melting of humidified fibers and gives structural information during the melting transition.

The fact that the evolution of the width is very similar for all the submerged samples and the humidified samples suggest that the confinement does not influence dramatically the correlation length along the molecule during the melting transition, although it has a significant effect in the melting temperature and the cooperativity as detected by calorimetry.

The PBD model was able to predict the melting curves of the submerged DNA by scaling the interaction between nucleotides by the melting temperature. A moderate variation of

the confinement does not seem to inhibit the successful application of the model. The results validate the use of the PBD model to describe complex DNA denaturation curves and includes structural information which is lacking for a complete understanding of the transition.

For Na-DNA fibers submerged in ethanol solutions calorimetry and X-ray were very consistent with the published literature regarding the B-to-A transition and the behavior of T_m and width of the melting. The room temperature study with neutrons detected the B-to-A transition at roughly the same ethanol concentration and proved that the salt concentration of the solution modifies the critical ethanol amount needed for the transition to occur. Neutron scattering also showed that for ethanol concentrations higher than 60% A/B contamination should be expected in Na-DNA fibers. Li-DNA were found to be free from B/A contamination no matter the ethanol concentration. The fibers swelled as the ethanol content decreased, i.e. the intermolecular distance decreased with ethanol concentration.

The fit of the temperature-dependent data of the ethanol samples showed a behavior very similar for I_0 and γ even though the fitting function for ethanol samples is clearly less optimal than for PEG samples. The study of the ethanol samples complement and expand the investigations performed on the PEG samples since the use of ethanol gives the chance to monitor how the changes in base stacking affect the correlations along the molecules. A complete investigation with further experiments, fitting procedures which better approximate the diffuse contribution of the ethanol samples; and a quantitative analysis using the PBD model will be perform in future works.

In chapter five, the room temperature study of the short chain Widom-601 sequence strongly suggested the presence of metastable kinks in the molecules in solution. This has not been observed before. The kinks could contribute to the strong histone positioning tendency that this sequence is famous for. The kink might favor the wrapping of the DNA to the histone octamer by mechanically preventing the diffusion of the octamer. These results can also reconcile the controversial reports about experimental measurements of DNA flexibility, which suggest there is a length-scale dependence of the flexibility of the DNA. For

short sequences, intrinsic curvature effects or kinks can make the experimentally detected persistence lengths appear shorter than for long molecules without implying that the shorter molecules are more flexible.

Regarding the temperature dependence scattering data of the Widom-601 sequence, the analysis qualitatively defends the hypothesis that persistence length and flexibility of DNA decrease as temperature is raised and that this decrease is correlated with the onset of the melting transition. Improving the accuracy of the analysis using fittings of the $S(Q)$ whit simple homogenous-polymer models looks unlikely. Combining the extended Kraty-Porod model presented sec. 5.1 with the conformational search method and small angle scattering data is a promising way to gain quantitative information on the temperature dependence of the flexibility of DNA.

Part of the work presented in this thesis has been included in two papers submitted for publication, the references of these papers are:

- A. González, A. Wildes, M. Marty-Roda, S. Cuesta-López, E. Mossou, A. Studer, B. Demé, Gaël Moiroux, J. Garden, N. Theodorakopoulos and M. Peyrard. *The melting of oriented DNA fibers submerged in polyethylene glycol solutions studied by neutron scattering and calorimetry.* Submitted to J. Phys. Chem. B.

- A. González, T Schindler, R. Boopathi, M. Marty-Roda, L. Romero-Santacreu, A. Wildes, L. Porcar, A. Martel, N. Theodorakopoulos, S. Cuesta-López, D. Angelov, T. Unruh and M. Peyrard. *Kinky DNA in solution? Small angle scattering study of a nucleosome positioning sequence. Submitted to NAR.*

Acknowledgements

I would like to thank the ILL and all its staff for the financial support and all the kind help received these three years; my supervisors who did an incredible job taking care of me and my project; Jean-Luc Garden and Gaël Moiroux for teaching me a lot about calorimetry; David Hess for lab support; and specially to Michel Peyrard for contributing greatly to the work presented in this thesis.

Bibliography

[1] M. Peyrard. Melting of the double helix, *Nature physics*, **2006**, 2, 13–14.

[2] S. Cuesta-López, D. Angelov, and M. Peyrard. Adding a New Dimension to DNA Melting Curves., *EPL*, **2009**, 87, 48009.

[3] H. Tateishi-Karimta and N. Sugimoto. Control of Stability and Structure of Nucleic Acids Using Cosolutes., *Methods*, **2014**, 67, 151–158.

[4] R.M. Wartell and A.S. Benight. Thermal Denaturation of DNA Molecules: A Comparison of Theory with Experiment., *Physics Reports*, **1985**, 126, 67–107.

[5] C.T. Wittwer. High-Resolution DNA Melting Analysis: Advancements and Limitations., *Hum. Mutat.*, **2009**, 6, 857–859.

[6] L. Garibyan and N. Avashia. Research Techniques Made Simple: Polymerase Chain Reaction (PCR), *J. Invest. Dermatol.*, **2013**, 133, doi:10.1038/jid.2013.1.

[7] A. Wildes, N. Theodorakopoulos, J. Valle-Orero, S. Cuesta-López, J.-L. Garden, and M. Peyrard. Structural Correlations and Melting of B-dna Fibers., ***Phy. rev. E***, **2011**, 83, 61923.

[8] J. Valle-Orero, A. Wildes, N. Theodorakopoulos, S. Cuesta-López, J-L Garden, S. Danikin, and M Peyrard. Thermal Denaturation of A-DNA., *New Journal of Physics*, **2014**, 16, 113017.

[9] M. Peyrard and A.R. Bishop. Statistical Mechanics of a Nonlinear Model for DNA Denaturation., *Phys. Rev. Lett.*, **1989**, 62, 2755–2758.

[10] S. Yasar, R. Podgornik, J. Valle-Orero, M.R. Johnson, and A. Parsegian. Continuity of States Between the Cholesteric-line Hexatic Transition and the Condensation Transition in DNA Solutions., *J. Phys. Chem. B*, **2013**, 117, 1849–1856.

[11] R Podgornik, H.H Stray, K. Gawrisch, D.C Rau, A. Rupprecht, and V.A. Parsegian. Bond Orientational Order, Molecular Motion, and Free Energy of High-density DNA Mesophases., *Proc. Natl. Acad. Sci. USA*, **1996**, 93, 4261–4266.

[12] A. Wildes, L. Khadeeva, T. William, J. Valle-Orero, A. Studer, J.-L. Garden, and M. Peyrard. Melting of Highly Oriented Fiber DNA Subjected to Osmotic Pressure., *J. Phys. Chem. B*, **2014**, 118, 3785–3792.

[13] CA. Davey D. Vasudevan, EYD. Chua. Crystal structures of nucleosome core particles containing the '601' strong positioming sequence., *Journal of Molecular Biology*, **2010**, 403, 1–10.

[14] JD. Kahn. Dna, flexibly flexible., *Biophysical Journal*, **2014**, 107, 282–284.

[15] Nikos Theodorakopoulos and Michel Peyrard. Base pair openings and temperature dependence of dna flexibility., *PRL*, **2012**, 108, 078104.

[16] A. Rupprecht. Preparation of oriented dna by wet spinning, *Acta chemica Scandinavica*, **1966**, 20, 494–504.

[17] A. Rupprecht. A wet spinning apparatus and auxiliary equipment suitable for prepating samples of oriented dna, *Biotechnology and Bioengineering*, **1970**, 12, 93–121.

[18] A. Rupprecht. Preparation of oriented dna in large amounts, *Biochemical and biophysical research communications*, **1963**, 2, 163–168.

[19] J. Widom PT. Lowary. New dna sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning, *Journal of Molecular Biology*, **1998**, 276, 19–42.

[20] H.J. Flammershein G. Hohne, W. Hemminger. *Differential Scanning Calorimetry. An introduction for Practitioners.* Springer (Ed), 1993.

[21] M. Randall G. N. Lewis. *Thermodynamics (2nd ed.).* McGraw-Hill Book Co., 1961.

[22] S. S. Zumdahl. *Thermochemistry.* Cengage Learning, 2008.

[23] J. Valle Orero. *Dynamics and thermal behaviour of films of oriented DNA fibres investigated using neutron scattering and calorimetry techniques.* École Normale Supérieure de Lyon, 1981.

[24] E. Cotts M. Patt, B. White. Thermal time constants in differential scanning calorimetry, *Thermochimica Acta*, **1992**, 197, 413–424.

[25] Liya Khadeeva. *Master thesis: Structural and Dynamical Aspects of Oriented DNA Fibers.* University of Rennes, University of Turin, Institut Laue Langevin, 2013.

[26] M. Peyrard J. Valle-Orero, J. Garden. Glassy behavior of denatured dna films studied by differential scanning calorimetry, *Physical Chemistry B*, **2012**, 116, 4394–4402.

[27] Setaram instrumentation. *micro DSC III. Commissioning utilisations.* Setaram instrumentation, 2006.

[28] A. L. Feig. *Methods in enzymology(Calorimetry).* Elsevier, 2016.

[29] C. Spink. Differential scanning calorimetry, *Methods in Cell Biology*, **2008**, 84, 115–141.

[30] C. Spink. The deconvolution of differential scanning calorimetry unfolding transitions, *Methods*, **2015**, 76, 78–86.

[31] M. Peyrard J. Valle-Orero, A. Wildes. Thermal denaturation of a-dna, *New journal of physics*, **2014**, 00, 00–00.

[32] D.S. Sivia. *Elementary scattering theory.* Oxford Unilever, 2011.

[33] G. L. Squires. *Introduction to the theory of thermal neutron scattering.* Cambridge University Press, 1978.

[34] A. Lucas and P. Lambin. Diffraction by dna, carbon nanotubes and other halical nanostrucutres., *Reports on Progress in Physics*, **2005**, 68, 1181–1249.

[35] M. Polanyi. The x-ray fibre diagram z., *Phys.*, **1921**, 7, 149.

[36] P. Ewald. Introduction to the dynamical theory of x-ray diffraction, *Acta Cryst. sect. A*, **1969**, 25, 103.

[37] W. Cochran, F. Crick, and V. Vand. The structure of synthetic polypeptides i: The transform of atoms on a helix., *Acta Cryst.*, **1952**, 5, 581.

[38] R. Langridge and H.R. Wilson. The Molecular Configuration of Deoxyribonucleic Acid.I, *J. Mol. Biol.*, **1960**, 2, 19–37.

[39] R. Franklin and R. Gosling. The structure of sodium thymonucleate fibers i: The influence of water content., *Act. Crys.*, **1953**, 6, 673.

[40] R. Franklin and R. Gosling. The structure of sodium thymonucleate fibers ii: The cylindrically symmetrical patterson function., *Act. Crys.*, **1953**, 76, 378.

[41] W. Fuller, T. Forsyth, and A. Mahendrasingam. Water-dna interactions as studied by x-ray and neutron fibre diffraction., *the Royal society.*, **2004**, –, –.

[42] W. Astbury. X-ray studies in nucleic acids., *Phil. Trans. R. Soc. Lond.*, **1931**, 230, 75–101.

[43] V.F. Sears. Neutron scattering lengths and cross sections., *Neutron news*, **1992**, 3:3, 26–37.

[44] T. Unruh T. Schindler, J. Walter. In situ study on the evolution of multimodal particle size distributions of zno quantum dots, *The journal of Physical Chemistry B*, **2015**, 119, 15370–15380.

[45] M. Peyrard A. Wildes, L. Khadeeva. Melting of highly oriented fiber dna subjected to osmotic pressure, *The journal of Physical Chemistry B*, **2015**, 119, 4441–4449.

[46] I. Grillo. *Effect of instrumental resolution and polydispersity on ideal form factos in small angle neutron scattering.* ILL technical report, 2001.

[47] D. Svergun. Determination of the regularisation parameter in indirect-transform methods using perceptual criteria., *Journal of applied crystallography*, **1992**, 25, 495–503.

[48] A. Tikhonov and V. Arsenin. *Solution of Ill-Posed Problems.* NY: Wiley, 1977.

[49] https://www.embl-hamburg.de/biosaxs/manuals/gnom.

[50] F. Zhang, J. Illavsky, G. Long, J. Quintana, A. Allen, and P. Jemian. Glassy carbon as an absolute intensiy calibration standard for small-angle scattering., *Mettall and Mat. Trans. A*, **2010**, 41, 1151–1158.

[51] J. Watson. *The double helix: a personal account of the discovery of the structure of DNA.* New york: Atheaeum, 1961.

[52] E.F. Flanklin and R.G. Gosling. The structure of sodium thymonucleate fibres. i the influence of water content, *Acta Cryst.*, **1953**, 6, 673.

[53] J.D. Watson and F.H. Crick. Molecular structure of nucleic acids, *Nature*, **1953**, 171, 737–738.

[54] M. Wilkins, A. Stokes, and H. Wilson. Molecular structure of deoxypentose nucleic acids, *Nature*, **1953**, 171, 737–738.

[55] W. Fuller, M.H. Wilkins, H.R. Wilson, and L.D. Hamilton. The molecular conformation of deoxyribonucleic acid: Iv, *Journal of Molecular Biology*, **1965**, 12, 60–80.

[56] R.E. Franklin and R.G. Gosling. The molecular conformation of deoxyribonucleic acid: I, *Journal of Molecular Biology*, **1960**, 2, 19–37.

[57] V. Forsyth, A. Mahendrasingam, P. Langan, W. Pigram, E. Stevens, and Y. Al-Hayalee. X-ray and neutron high-angle fibre diffraction studies in the determi- nation of hydration and ionic structure around dna., *Neu. and Xr, scatt. compl. techn.*, **1990**, 12, 237–248.

[58] S. Lindsay, S. Lee, and J. Powell. The origin of the a to b transition in dna fibers and films., *Biopolymers*, **1988**, 27, 1015–1043.

[59] F. Crick and J. Kendew. X-ray analysis and protein and protein structure., *Adv. Protein Chem*, **1957**, 12, 133–213.

[60] J. Valle-Orero, A. Wildes, J. Garden, and M. Peyrard. Purification of A-form DNA fiber samples by the removal of B-form dna residues., *J. Phys. Chem. B*, **2013**, 117, 1849–1856.

[61] S. Zimmerman and B. Pheiffer. A direct demonstration that the ethanol-induced transition of dna is between the a and b forms: an x-ray diffraction study., *J. Mol. Biol.*, **1979**, 135, 1023–1027.

[62] S. Zimmerman and B. Pheiffer. The molecular configuration of deoxyribonucleic acid i. x-ray diffraction study of a crystalline form of the lithium salt., *J. Mol. Biol.*, **1979**, 142, 315–330.

[63] D.B. Knowles, A.S. LaCroix, N.F. Deines, L. Shkel, and M.T. Record. Separation of Peferential Interaction and Excluded Volume Effects on DNA Duplex and Hairpin Stability., *Proc. Natl. Acad. Sci.*, **2011**, 108, 12699–12704.

[64] H. Karimata, S. Nakano, and N. Sugimoto. Effect of polyethylene glycol on dna duplex stability at different nacl concentrations., *Bull. Chem. Soc. Jpn.*, **2007**, 10, 1987–1994.

[65] R. Politi and D. Harries. Enthalpically driven peptide stabilization by protective osmolytes.., *Chem. Commun*, **2010**, 46, 6449–6451.

[66] A. Badasyan, S. Tonoyan, A. Giacometti, R. Podgornik, V.A. Parsegian, Y. Mamasakhlisov, and V. Morozov. Osmotic Pressure Induced Coupling between Cooperativity and Stability of a Helix-Coil Transition., *PRL*, **2012**, 109, 068101.

[67] D Rau, B. Lee, and V. Parsegian. Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel dna double helices., *Proc. Natl. Acad. Sci. USA*, **81**, 9, 2621–2625.

[68] A. Kornyshev and S. Leikin. Theory of interaction between helical molecules., *J, Chem. Phy.*, **1997**, 107, 3656.

[69] A.G. Cherstvy and A.A. Kornyshev. DNA Melting in Aggregates: Impeded or Facilitated., *J. Phys. Chem. B.*, **2005**, 109, 13024–13029.

[70] Parsegian5. Hydration forces., *Annu. Rev. Phys. Chem.*, **1993**, 44, 369–95.

[71] A. Kornyshev and S. Leikin. Electrostatic zipper motif for dna aggregation., *PRL*, **1999**, 82, 4138.

[72] R. Owczarzy. Melting temperatures of nucleic acids: Discrepancies in analysis., *Biophysical chemistry*, **2005**, 117, 207–215.

[73] C. H. Spink and J. B. Chaires. Selective stabilization of triplex dna by poly(ethylene glycols)., *Journal of the american chemical society*, **1995**, 117, 12887–12888.

[74] Effect of osmoregulatory solutes on the thermal stability of calf-thymus DNA. Barone, g. and del vecchio, p. and esposito, d. and fessas, d. and grazia, g., *J. Chem. Soc. Faraday Trans.*, **1996**, 92, 1361–1368.

[75] C.H. Spink. Differential Scanning Calorimetry, *Meth. Cell Biol.*, **2008**, 84, 115–141.

[76] A. Vedenov, A. Dykhne, and D. Frank-Kamenerskii. The helix-coil transition in dna., *Sov. Phy. Usp.*, **1972**, 14, 715.

[77] A. Cherstvy, A. Kornyshev, and S. Leikin. Torsional deformation of double helix in interaction and aggregation of dna.., *J. Phys. Chem. B*, **2004**, 108, 6508–6518.

[78] A. Cherstvy. *Interaction, Recognition and Condensation of DNA Duplexes.* Thesis. Dusseldorf University., 2002.

[79] W Kabsch, C. Sander, and E. Trifonov. The ten helical twist angles of b-dna, *Nucl. Acid. Res.*, **1982**, 10, 1097–104.

[80] V. Gorin, A. Zhurkin and W. Olson. B-dna twisting correlates with base-pair morphology., *J. Mol. Biol.*, **1995**, 247, 34–48.

[81] J. Schutt. Effect of anisotropic bending rigidity and finite twisting rigidity on statistical properties of dna model filaments.., *Biopolymers*, **1985**, 24, 1215–1232.

[82] C. Spink and J. Chaires. Effects of Hydration, Ion Release, and Excluded Volume on the Melting of Triplex and Duplex DNA., *Biochemistry*, **1999**, 38, 496–508.

[83] P. Woolley and P.R. Wills. Excluded-volume Effect of Inert Macromolecules on the Melting of Nucleic Acids., *Biophysical Chemistry*, **1985**, 22, 89–94.

[84] G. Baldini, H. Fu-Hua, G. Varini, L. Cordone, S. Fornilim, and G. Onori. Dna melting induced by alcohols: Role of the solvent properties., *Nuovo Cimento*, **1985**, 6, 618–630.

[85] A. Rupprecht, J. Piskur, J. Schultz, Nordenskioldm L., Z. Song, and L. Gojmir. Mechanochemical study of conformational transitions and melting of li-, na-, k-, and csdna fibers in ethanol-water solutions., *Biopolymers*, **1994**, 34, 897.

[86] J. Piskur and A. Rupprecht. Aggregated dna in ethanol solution., *FEBS letters*, **1995**, 375, 174–178.

[87] A. Rupprecht and J. Piskur. A simple mechanochemical method for studying structure and dynamics pf biopolymer fibers in varios media., *Acta Chem. Scand.*, **1983**, 9, –.

[88] T. Weidlich, S. Linsay, and A. Rupprecht. Counterion effect on the structure and dynamics of solid dna., *PRL*, **1988**, 61, 1674–1677.

[89] M. Zehfus and J. Curtis. Conformation of p-form dna., *Biopolymers*, **1978**, 17, 523–525.

[90] T. Dauxois, M. Peyrard, and A. R. Bishop. Entropy-driven dna denaturation., *Physical review E*, **1993**, 47, R44–47.

[91] M. Peyrard and A. R. Bishop. Statistical mechanics of a nonlinear model for dna denaturation., *Physical review letters*, **1989**, 62, 2755–2758.

[92] N. Theodorakopoulos. Melting of Genomic DNA: Predictive Modeling by nonlinear lattice dynamics., *Phys. Rev. E*, **2010**, 82, 021905.

[93] T. Dauxois and M. Peyrard. Entropy-Driven DNA Denaturation., *Phys. Rev. E*, **1992**, 47, 47R44.

[94] A. Wildes, J. Theodorakopoulos, Valle-Orero, S. Cuesta-López, J.-L. Garden, and M. Peyrard. Thermal Denaturation of DNA Studied with Neutron Scattering., *PRL*, **2011**, 106, 048101.

[95] M. Krisch, A. Mermet, H. Grimm, V.T. Forsyth, and A. Rupprecht. Phonon Dispersion of Oriented DNA by Inelastic X-ray Scattering, *PRE E*, **2006**, 73, 061909.

[96] O. Kratky and G. Porod. Rontgenunterschung geloster fadenmolekule, *Recl. Trav. Chim. Pays Bas*, **1949**, 68, 1106–1122.

[97] W. Fuller and M. Wilkins. The molecular configuration of deoxyribonucleic acid iv. x-ray diffraction study of the a form., *J. Mol. Biol.*, **1965**, 12, 60–80.

[98] R. Langridge, H. Wilson, C.W. Hooper, M. Wilkins, and L. Hamilton. The molecular configuration of deoxyribonucleic acid i. x-ray diffraction study of a crystalline form of the lithium salt., *J. Mol. Biol.*, **1965**, 12, 60–80.

[99] D. Grasso, S. Fasone, and C. La Rosa. A Calorimetruc Study of the Different Thermal Behaviour of the DNA in the Isotropic and Liquid-Crystalline States., *Liq. Cryst.*, **1991**, 9, 299–305.

[100] F. Sebastine, A. Pietrini, M. Longo, L. Comez, C. Petrillo, F. Sacchetti, and A. Paciaroni. Melting of dna nonoriented fibers: A wide-angle x-ray diffraction study., *J. Phys. Chem. B*, **2014**, 118, 3785–3792.

[101] W. Russel, D. Saville, and W. Schowalter. *Colloidal Dispersions.* Cambridge University press., 1989.

[102] R. Owczarzy. Melting tempeature of nucleic acids: Discrepancies in analysis., *Biophysical Chemistry*, **2005**, 177, 207–215.

[103] R. Wells, J. Larson, and R. Grant. Physicochemical studies on polydeoxyribonucleotides containing defined repeating nucleotide sequences., *J. Mol. Bio.*, **1970**, 54, 465–497.

[104] L. van Eijck, F. Merzel, S. Rols, J. Ollivier, V.T. Forsyth, and M.R. Johnson. Direct Determination of the Base-Pair Force Constant of DNA from acoustic Phonon Dispersion of the Double Helix., *PRL*, **2011**, 107, 088102.

[105] N.L. Quinn, W. Levenkova, N. Chow, P. Bouffard, K.A. Boroevich, J.R. Knight, T.P. Jarvie, K.P. Lubieniecki, B.A. Desany, B.F. Koop, T.T. Harkins, and W.S. Davidson. Assesing the Feasibility of GS FLX Pyrosequencing for Sequencing the Atlantic Salmon genome., *BMC Genomics*, **2008**, 9, 404–418.

[106] R. Lavery, M Moakher, J.H. Maddocks, D. Petkeviciute, and K. Zakrewska. Conformational Analysis of Nucleic Acid Revisited: Curves+., *Nucl. Acid Res.*, **2009**, 37, 5917–5929.

[107] JD. Kahn. Dna, flexibly flexible., *Biophysical Journal*, **2014**, 107, 282–284.

[108] T.E. Cloutier and J. Widom. Spontaneous sharp bending of double stranded dna, *Molecular Cell*, **2004**, 14, 355–362.

[109] R. Vafabakhsh. Extreme bendability of dna less than 100 base pairs long revealed by single-molecule cyclization, *Proc. Natl. Acad. Sci. USA*, **2012**, 102, 5397–5402.

[110] R.S. Mathew-Fenn, D. Rhiju, and A.B. Harbury. Remeasuring the double helix, *Science*, **2008**, 322, 446–449.

[111] F. Crick and A. Klug. Kinky helix., *Nature*, **1975**, 255, 530–533.

[112] M.S. Ong, T.J. Richmond, and C.A. Davey. Dna streching and extreme kinking in the nucleosome core., *J. Mol. Biol.*, **2007**, 368, 10671074.

[113] D. Svergun. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing., *Biophys. J.*, **1999**, 76, 2879–2886.

[114] F. Lankas, R. Lavery, and J. Maddocks. Kinking occurs during molecular dynamics simulations of small dna minicircles., *Structure*, **2006**, 14, 1527–1534.

[115] EYD Chua, D. Vasudevan, E. Gabriela, B.W. Davey, and CA Davey. The mechanics behind dna sequence-dependent properties of the nucleosome., *Nucl. Acids Res.*, **2012**, 40, 6338–6352.

[116] JL. Mergny and L. Lacroix. Analysis of thermal melting curves, *oligonucleotides*, **2003**, 13, 515–537.

[117] C. Von Sonntag. *The Chemical Basis of Radiation Biology.* Taylor and Francis, London, 1989.

[118] Bednar. J., P. Furrer, V. Katrich, A. Stasiak, J. Dubochet, and A. Stasiak. Determination of the dna persistence lenght by cryo-electron microscopy. separation of the static and dynamic contributions to the apparent persistence length of dna., *J.Mol.Biol.*, **1995**, 254, 579–594.

[119] H.B. Gray and J.E. Hearst. Flexibility of native dna from the sedimentation behavior as a function of molecular weight and temperature., *Journal of Molecular Biology*, **1968**, 35, 111–129.

[120] S. Geggier, A. Kotlyar, and A. Vologodskii. Temperature dependece of dna persistence length., *Nucleic Acids Research*, **2011**, 39, 1419–1426.

[121] G. Weber, J.W. Essex, and C. Neylon. Probing the microscopic flexibility of dna from melting temperatures., *Nature Phys*, **2009**, 5, 769.

[122] www.sasview.org.

[123] J.S. Pedersen and P. Schurtenberger. Scattering functions of semiflexible polymers with and without excluded volume effects., *Macromolacules*, **1996**, 29, 7602.

[124] G Jerke, J.S. Pedersen, S.U. Egelhaaf, and P. Schurtenberger. Static structure factor of polymer like micelles: Overall dimension, flexibility and local properties of leticin reverse micelles in deuterated isooctanes., *Physical Review E*, **1997**, 56, 5772.

[125] W. Chen, P.D. Butler, and J. Magid. Incorporating intermicellar interactions in the fitting of sans data from cationic wormlike micelles., *Langmuir*, **2006**, 22, 6539.

[126] https://www.ncnr.nist.gov/resources/activation/.

[127] R. C. Johan and Van der Maarel. Introduction to biopolymer physics. Chapter 2,pag 31.