A Multidisciplinary approach to DNA Nano/Bio-Physics

UNIVERSIDAD
DE BURGOS

Marta Marty-Roda

Supervisor: Dr. Santiago Cuesta-López

Advanced Materials, Nuclear Technology, Applied Nanotechnology
University of Burgos

This dissertation is submitted for the degree of
Doctor of Chemistry

ICCRAM March 2018

A mis padres, por confiar siempre en mí y enseñarme a superarme siempre ante cualquier adversidad . . .

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except when specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Marta Marty-Roda
March 2018

Acknowledgements

El refranero español tiene un refrán que viene muy a cuento y que dice así: "*Es de bien nacidos, ser agradecido*" y yo realmente tengo mucho que agradecer. Han sido muchas las personas que han influido de una manera u otra a lo largo de mi formación académica y personal, y que en alguna medida son responsables de la presente tesis doctoral, pero ante la imposibilidad de nombrar a todos, mencionaré a aquellos que forman parte de mi presente y casi con total seguridad estarán presentes en mi futuro.

En primer lugar, a mi director de tesis, Santiago Cuesta, quien me dió esta oportunidad confiando en mi desde el primer momento para este proyecto y que me ha ilustrado durante este proceso con su experiencia y conocimientos, enderezando mi rumbo y proporcionando la ayuda necesaria siempre que me ha hecho falta. Además ha ayudado a promover en mí el desarrollo de un sinfín de habilidades tales como independencia, motivación, multidisciplinaridad y sinergia muy valiosas en el mundo científico y en el personal.

También me gustaría dar las gracias a Roberto Serrano y Lorena Romero. A Roberto, por ser el primero en ofrecerme la oportunidad en aquella entrevista por Skype, y por ser una inagotable fuente de apoyo y ayuda. No importa cuán difícil sea lo que le pidas, que lo intenta resolver al momento y encima con una sonrisa en la cara. Su excelente disposición es muy de agradecer y su sonrisa contagia y alegra hasta el más gris de los días. Gracias. Y a Lorena por toda la ayuda y la orientación en la parte experimental, enseñándome a no rendirme ante las adversidades y que la constancia y los esfuerzos realizados siempre tienen recompensa, estando siempre dispuesta a escucharme y aconsejarme. Gracias por todas las palabras de aliento y no dejarme desfallecer, y por tener siempre la solución a los problemas que iban surgiendo.

I have had the great opportunity of complete my thesis with some internships in Norway and France so I must also thanks the people who did it possible. They did everything in their power to make my stay in their places easier, and their enthusiasm and knowledge were vital for me to never return with my hands empty of my stays outside Burgos. To professor Titus Sebastiaan van Erp for all the great help and guidance throughout my stays at NTNU of Norway. I really appreciate the help he provided me with for understanding the theory behind his work, dealing with problems related to the simulations, evaluating the

results obtained and much more. I also want to thanks Oda Dahlen for making my days at Trondheim unforgettable. To Andrew Wildes and Hervé Guillou for to invite me to Institut Laue-Langevin and Institut NEEL/CNRS respectively. To Andrew for his enthusiasm and happines and for making the hard days and nigths doing experiments in the instrument of ILL more enjoyable. Also to Herve for making my days at Grenoble easier and bringing me into the AFM world. During this years I have had the big chance to meet very experienced people in different fields with different ways of working, and I would like to thank them for their ideas, all the valuable, helpful and instructive discussions,and their time and effort: Michael Peyrard, Nikos Theodorakopoulos, Jean-Luc Garden and Enrico Riccardi.

Y a Adrián González por tantas noches de experimentos compartidas tanto en Grenoble como en Burgos. Por amenizar esas noches con sus charlas y su visión de la vida y porque haya aguantado en el mundo científico a pesar de ellas. Gracias por esnseñarme tanto sobre neutrones y transición de ADN en fibra. Para todo, PhD twins.

En un enfoque más cotidiano y personal me gustaría agradecer a gente que ha sido participe de una manera menos académica en este proceso: A Rodri, por hacer desde el primer momento mi estancia en Burgos especial y diferente. Por enseñarme la ciudad, compartir conmigo sus lugares y enseñarme a ver las cosas con una perspectiva distinta, aunque nunca acabe de darme la razón. A la gente que forma parte del ICCRAM, por compartir el día a día, los cafés, las horas de trabajo, etc, pero especialmente a Marta, Claudia y Noemí, somos las iccramers que más lo petan! Gracias por cada charla, cada cerveza, canción de las grecas, cada beer pong, whatsapp de ánimo, por compartir cada locura inventada, cada consejo o cada hora de deporte... Además, gracias Mart por leerte este tostón y todos los valiosos comentarios al respecto; Noemí por ser la ayudante de maquetación oficial y el apoyo mostrado ante cada duda, además de todas las horas de charla y elucubraciones (aunque sea a gritos jaja); y Clau, por hacerme reir cada mañana con tu bendita locura (estás como la puerros)... Gracias también a Sandra, Bea, Noel, Sergio, Roci, Gloria, Sonia, Jaime, Seba, Pablo, Juanan, Carlos, Brixhilda, Sergiu, Iris, Eliosa, Mari Luz... en fin todos los iccramers! Habéis hecho estos duros últimos meses mucho más llevaderos... Entre todos habeís sido un verdadero apoyo y la familia que no tengo en Burgos... gracias por tanto, y sé que siempre os tendré ahí ... Gracias por todos los buenos momentos compartidos (y los no tan buenos), las cañas y los vinos que surgen con cualquier excusa y los super desayunos... Queda pendiente el Internacional Culinary Center!

Mi verdadera familia, mis padres y mis hermanos, tíos y primos, por todo su apoyo, amor y ayuda durante estos cuatro años. Por creer en mí, por festejar mis pequeños logros y compartirlos con el resto del mundo orgullosos de que los haya conseguido. Por inculcarme los valores del esfuerzo, el respeto y el sacrificio desde el principio. Por confiar en que

siempre puedo lograr lo que me pronpongo. Y nunca dudar de mí. Por hacerme ser la persona que hoy día soy, y por no dejarme rendirme nunca y no rendirse ellos tampoco.

Y al resto de personas que han influido durante todo este proceso, incluso antes de que lo iniciara. A Carmiña, por inculcarme el amor por la ciencia durante los 4 años que ha sido mi profesora de biología. Por despertar en mí la necesidad de saciar mi curiosidad y de no conformarme si no sé el por qué de las cosas. Y mis amigos, los de toda la vida, a los que compartieron los años de Química conmigo, y los que he ido haciendo durante estos cuatro años. A Almu, mi rubia, mi zipi, mi bob... sabes que eres mi mitad y que sin ti, sin tu apoyo y tus frases adecuadas en los momentos precisos nada hubiera sido igual. Gracias por compartir conmigo mucho más que la química. A Vane, Maritere, Lore, Naira por estar siempre ahí, por compartir tanto, desde mis inquietudes hasta mis logros y siempre estar dispuestas y hacerme vivir momentos memorables. Gracias a todas por vuestro incondicional apoyo y paciencia. Y al resto de gente con la que he compartido tanto en la facultad... Danielo, madridista y entrenador, aunque nos vemos menos de lo que debiéramos siempre nos tendremos de apoyo. Y Sandra, Tara, Elisa... con las que he compartido desde vestuario en el futbol hasta viajes inolvidables.

Y por último, pero no menos importante, a mis BCP... sabeís que Pinar es una parte imprescindible de mi y vosotras sois mi Pinar... No me imagino mis veranos sin estar en "*La Poza*" rodeada de vosotras. Y que no importa los kilométros que el destino se empeñe en separarnos, que sé que siempre puedo contar con vosotras! Y ahora más si cabe con los juntaitos. Gracias por todo y por estar desde siempre y para siempre...

Este trabajo ha sido posible gracias a los proyectos del Ministerio de Economía, Industria y Competitividad-FIS2012-38827- y de la Junta de Castilla y León-Nanofibersafe BU079U16-, especialmente gracias a la beca de Formación de Personal Investigador (BES-2013-065453) concedida en el marco del primer proyecto. Deseo agradecer igualmente a la Universidad de Burgos, el fondo Anders Jahre (Project 40105000) y nuevamente al MICINN (EEBB-I-2015-09973) por la financiación de mis estancias en Noruega y Francia.

Abstract

DNA is an essential molecule for life, since it is the carrier of genetic information. To be able to transmit and to read the information encoded in the DNA, processes such as replication or gene transcription are necessary. These processes start with the local opening and the formation of bubbles in the DNA double helix. The structure of DNA has been widely studied and is really well known. However, the understanding of DNA dynamics, which determinantly influences the biological processes, is limited. Our effort in this memory will be dedicated to study these dynamics, concretly during the thermal denaturation of the DNA molecule.

The base pairs that form the DNA molecule can fluctuate even at biological temperature, producing a break in the bases and a local separation of the strands, known as denaturation bubbles. As the temperature increases, these bubbles grow, since more base pairs are opened, until the two strands are completely separated. This process is known as the melting transition of DNA.

In this work, the denaturation of DNA has been explored by different theoretical and experimental techniques. Although historically this phenomenon has been studied mainly in solution, in this thesis we will approach its study with ordered samples used more in structural than dynamic studies.

A theoretical study was carried out through simulations of molecular dynamics applying the Peyrard-Bishop-Dauixous (PBD) model. An implementation of this model has been developed for the study and understanding of biomolecules as DNA hairpins, also involved in important biological processes. Based on previous studies carried out by the group, we have added an entropic barrier to the Morse potential of the original PBD model, which takes into account the dependence of the stacking of the dimer base pairs. With this modification we will study the opening constant of DNA hairpins and we will compare the values obtained with experimental data to verify the validity of the improvement. This model has been applied to dsDNA chains to study their rates but also the size and distribution of bubbles in these sequences.

The preparation of samples and their characterization are also a fundamental part of this research. We explain the methodology followed to obtain the samples used in the experiments

that have been carried out. DNA films formed by oriented fibers will be used to study the melting transition with both neutron scattering and Raman spectroscopy. For the other samples, a specific synthesis methodology has been developed for the design of small DNA fragments, with controlled sequence. With them, the flexibility of DNA has been studied when performing biological functions. The role of the sequence and consequently of the DNA bubbles in specific situations of biological interest has been interpreted. Moreover, the persistence length of DNA has been determined when it is wound up for packaging around nucleosomes.

DNA fibers have also been widely used throughout this work. The obtaining of these highly ordered samples is explained in chapter 5 as well as their use to study changes in the correlation length during the melting transition correlating these changes with structural changes produced in the denaturation. This experiment has been carried out by means of neutron scattering, but also by Raman spectroscopy studying the denaturation of DNA by analyzing the changes in the vibration modes of the base pairs of the DNA helix.

In addition to study of the fusion of the fibers (following changes in intensity and shift of the Raman frequencies) we have also analyze the classical denaturation of the the DNA in solution. In this regard, we performed a comparison of the obtained data to detect the possible differences between the two processes. Finally, the influence of spatial confinement has been investigated by comparing the Raman bands of the dry DNA fibers and fibers submerged in different solutions.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 DNA molecule structure

DNA (deoxyribonucleic acid) is the carrier of our genetic information and therefore responsible for maintaining life. The obtention of new knowledge about the DNA and it functioning allows us to understand several biological processes; for instance, both DNA replication and gene transcription. The latter begins with the unwinding of the double helix and the formation of specific DNA bubbles. Therefore, the knowledge of the DNA structure is not sufficient to understand the biologic processes [12], it is necessary to understand how this structure affects the equilibrium properties and the dynamic of the DNA molecule [10]. Despite the fact that a half century has passed since the discovery of the structure of the DNA double helix [13], major questions still remain regarding its thermodynamic behavior and stability [14]. The dynamics of DNA also plays a major role in the function of DNA, however, while the DNA structure is throughly examined and described, the knowledge of DNA dynamics is rather limited.

DNA molecule has a very well known structure [13]: it is composed of two chains which are twisted around each other. Each of the strands forming the double helix is formed by a frame known as backbone of DNA which is composed of sugars (deoxyribose) and the phosphate groups which are joined by a ester bond. The sugars are linked by a covalent bond (N-glycosidic bond) to one of the four possible types of nitrogenous bases that appear in the molecule : adenine (A), guanine (G), thymine (T) and cytosine (C) (Fig. 1.1). These two strands twist forming a double helix connected by the nitrogen base pairs. One of the biggest achievements of Watson and Crick [13] was determining the nitrogen bases to pair: the adenine of one single chain can join through two hydrogen bonds (dotted lines in figure 1.1) with a thymine of the opposite chain forming a base pair between complementary bases. In turn, guanine can only bind to a cytosine, but it does so by means of three hydrogen bonds

instead of two [15]. Thus, there are only two possible base pairs: adenine-thymine and guanine-cytosine and DNA has a complementary structure containing the same information in each strand [16].

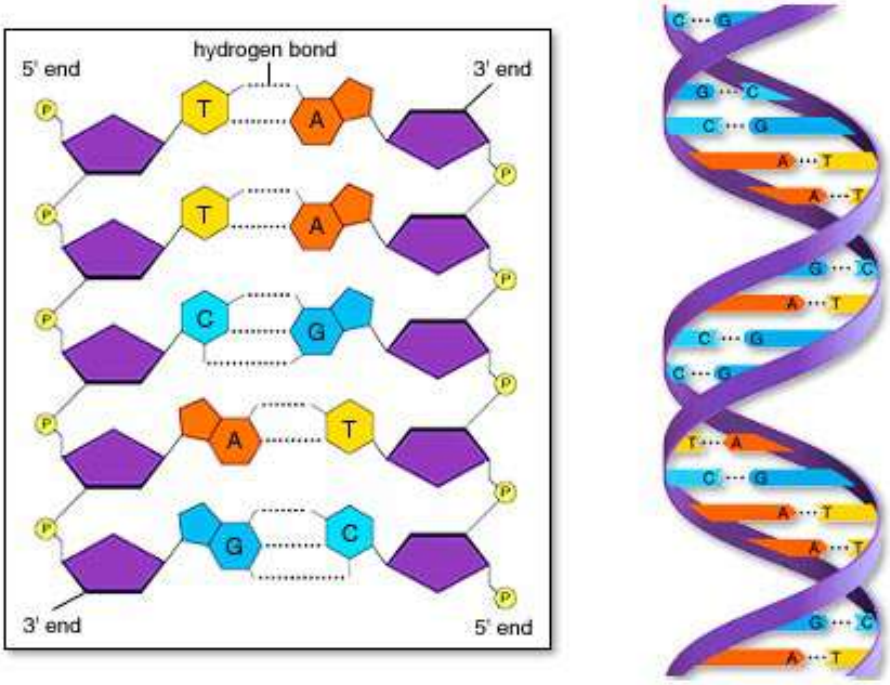


Fig. 1.1 Representation of primary (left) and secondary (right) structure of DNA. Phosphodiester bonds (yellow circles), sugar molecules (purple pentagons), nitrogenous bases (tagged rings) and hydrogen bonds (dotted line). The right panel shows the double helix structure

While the primary structure of DNA is determined by the sequence which store the genetic information and could be easily described by writing the abbreviation of bases, the secondary structure corresponds to the double helix structure. The backbone presents certain flexibility due to the rotation around the single bonds which is quite easy, while the helicoidal configuration restricts these rotations [17, 18]. The double helix of DNA is held together and stabilized by two main forces, hydrogen bonds and stacking interaction between base pairs, but DNA is not the static structure that diffraction pictures show [19, 11, 20]. These forces reflect the dynamics of the DNA molecule: while interaction between complementary bases is weak, the stacking interaction between the sugar-phosphate backbone due to hydrophobic forces and attractive interactions between aromatic rings as a result of overlapping of $\pi -$ *electrons* of the bases plateau is strong. This phenomena allows fluctuations of the molecule at biological temperature that result in the local opening of DNA which it is known as 'breathing of DNA' [21, 22]. The possibility of local opening is biologically relevant since any genetic process is initiated by the exposure of the base pairs sequence to the surroundings.

The transcription and replication of DNA, two central and vital processes for sustaining life, involve a dynamical alteration of the molecule and cannot take place without the unwinding of the DNA double helix and the formation of specific DNA bubbles. Further rising up the temperature after the formation of the afore mentioned unwinding regions, called "bubbles", extend them over the full molecule, resulting in a complete separation of the two strands. This is known as melting transition or thermal denaturation of DNA molecule.

Understanding the simple process of how the base pairs separate under thermal fluctuations is key in many biotechnology applications, such as the design of Polymerase Chain Reaction (PCR) probes or the HRM studies for biology, or molecular beacons (DNA hairpins). DNA denaturation is also becoming important for nanotechnology as DNA is now used, due to its self-assembly properties to create nanodevices or to design molecular memories [23, 24]. Appart from the biological relevance, DNA denaturation is a very interesting process by itself since involving a phase transition in a essentially 1D system [20, 21].

DNA thermal stability has been investigated in detail but it is also the main goal of this project. DNA molecule is the linker in each work presented in this thesis and the study of DNA denaturation has a main role. The second chapter reviews theoretical models which can be used to very simply modeling the thermochemical denaturation of the DNA molecule. After this, chapter three presents a brief synopsis of one of the most interesting architectures of DNA such as DNA hairpins, and uses what was studied in the previous chapter, mesoscopic models of DNA denaturation and DNA hairpins in order to study the dynamics and the rate constant of these molecules. Chapter four provides a description of the methodology used to prepare DNA samples which has been used to carry out all the experiments presented in this thesis. There is two main types of samples : DNA in solution but with a specific well-known sequence, which will be use in the study of DNA flexibility, and DNA oriented samples such as DNA films which has the purpose of studying melting transition with both Raman spectroscopy and Neutron scattering. Finally, chapter five exposes the investigation of DNA denaturation studied by Raman spectroscopy including a description of this experimental technique.

Chapter 2

Theoretical mesoscale DNA models

2.1 DNA thermal denaturation

The process in which double stranded DNA separates into two single strands by rising up the temperature is known as DNA melting transition or DNA thermal denaturation. We have already mentioned how important is to understand DNA melting since most of biological processes start with the opening of DNA base-pairs. Appart from temperature, denaturation process is also influenced by physical factors such as salt concentration or pH, and by the chain itself, being important the chain length or its specific sequence (AT/GC ratio). All this factors could influence the melting temperature (T_m), a parameter that characterizes the denaturation and which is defined as the temperature at which half of the melting has taken place. The denaturation process can be study by melting curves. The use of those curves is the canonical way to study denaturation transition and they lead to obtain thermodinamic properties of DNA, including ΔH, ΔS and ΔG. An example of denaturation curve is given in figure 2.1

In addition to its biological importance, DNA melting transition process also attracts the interest of physicists since it takes place by means of an entropy-induced phase transition. It is an example of a phase transition in a essentially 1D system where the entropy gained on passing from double-stranded DNA (dsDNA), which is more rigid than single-stranded DNA (ssDNA, much more flexible), may be enough to compensate the energy cost of breaking a base-pair, even at low temperatures. Both, entropic effects and the heterogeneity of the sequence, are involved in the energetic demand of the melting transition. Due to the strong dependence on all this factors, the dynamics of DNA molecule must be described by a nonlinear system.

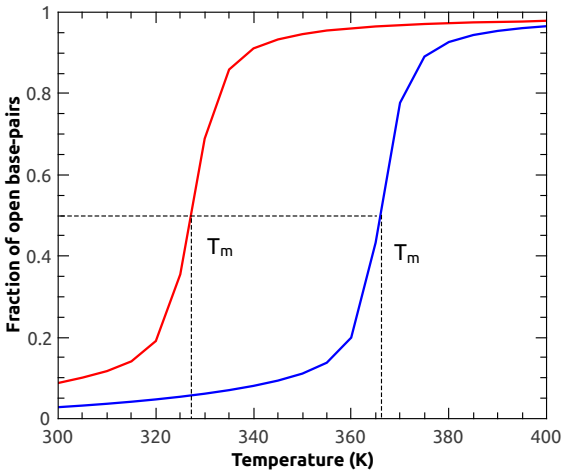


Fig. 2.1 Melting curves example for homogeneous chains of AT (red) and GC (blue) sequences. DNA length 200 base-pairs. The melting temperature is marked in each case.

2.2 DNA melting models

Historically, in the theoretical study of the denaturation transition various models have been used, since it has attracted the attention of theoreticians and experimental techniques have to be complemented. One problem is that melting transition, which is a dynamical process that can be compared with experiments, require very long timescales that are not reachable with full-atom simulations. The number of atoms that are needed to describe the DNA molecule is very large and therefore has a extremely large computational cost. This has created the necessity of mesoscopic theoretical models that describe the molecule as simple as possible and allow to study long DNA sequences [21]. The theoretical study of the transition of denaturation have used models based on Ising's model [25] which give different values based on formed or broken pairs; models that claim a thermodynamic description, as for example nearest-neighbor models [26]; Models of Poland-Scheraga [27–29]; simple zipper models [30, 31], or models that introduce a phenomenological potential between the bases to describe the hydrogen bonds like Peyrard-Bishop-Dauxois (PBD) model [12, 11, 32]. We introduce some of these models in this section.

2.2.1 Ising Model

Wilhelm Lenz (1920) and Ernst Ising (1925) introduced a model to study the behavior of ferromagnetic materials and to analyze the phase transition of the system [25, 33]. In this model they gave a value to a spin of the nucleus depending on its orientation (direction

of polarization). Doing an analogy with a DNA chain it is possible to describe the DNA denaturation using this Ising model. In this scenario, DNA is described as a sequence of two-state systems where the base pairs can be open, assigning them a value of 1, or close, being value 0, similar to Ising model with spin up and down. This model has created a precedent and consecutive models that try to describe the melting of DNA with a two-state system are known as Ising-like models.

In this model the size of helical parts (closed regions) is a parameter that can be described by a a critical exponent in a continuous phase transition, but the model does not take into account the form of opened regions (the entropy of the formed loops) or the cooperative effect of these regions. Thus, this approach cannot reproduce the full dynamics of the molecule. Ising models are well suited for long DNA chains and due to their simplicity are really convenient for practical calculations, but it is necessary to introduce non-linear dynamics in order to a proper description of the intermediate states.

2.2.2 Poland and Scherga Model

The Poland-Scherga model is a discrete model that was introduced in 1966 by Poland and Scherga [27, 28]. This model describes the DNA molecule like an alternating sequence of ordered (dsDNA) and unordered (ssDNA) states called *chain*. The dsDNA segments are modeled by paths while single strand DNA segments are modeled by loops. Here the bases could be in three states, bounded in the helix, free in loops or in unbound sequences between helicoidal parts. Figure 2.2 shows a schematic representation of the model. The bounded state is energetically favoured but, since the bases are weel-organized, it has no entropy. Meanwhile, the loop state has a high energy, but also carries a lot of entropy [34, 35].

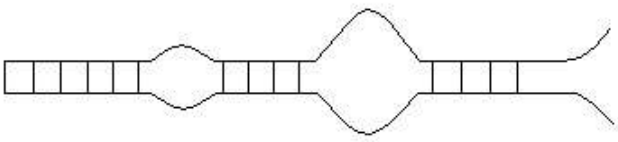


Fig. 2.2 Example of a schematic representation of the Poland-Scherga model.

Both, bound strand and loops of length l have a statistical weight given by the change in the entropy. The probability weight has the general form:

$$L(l) = \frac{As^l}{l^c} \tag{2.1}$$

where A and s are constants and the exponent c, which has been found that is $c = d/2$ (being d the dimension), has a crucial value. For $c \leq 1$ there is no phase transition and a first order transition arise if $c > 2$.

2.2.3 PBD Model

This is a more rigorous model to study the melting transition of the DNA that was introduced by Peyrard and Bishop in 1989 [12] and was improved with Dauxois in 1993 [11, 36]. The PBD is a coarse-grained, mesoscopic and dynamic model of the DNA molecule, which ignores its helicoidal structure, describes it as a connected chain of one-dimensional particles, and condenses all the atomic coordinates of a base pair into a single number y which describes the stretching of the bonds between the two bases. The simplest form of the model reduces DNA to N base pairs, each one of which is represented as a point-mass, m, along the DNA chain. These N one-dimensional particles represent the separation of a specific base pair from the grounding-state [19, 21]. The model solves the dynamical properties associated to the formation and stability of the bubbles providing reasonable statistics [37], succesfully predicts the persistence length and DNA flexibility [38] and reproduces the sharp phase transition seen in the melting curves of long homopolymers [20]. The phase transition is facilitated by an entropy induced process where the entropy gained when the system passes from the closed state to a denatured one, is enough to compensate the energy effort employed in the brakage of a base-pair.

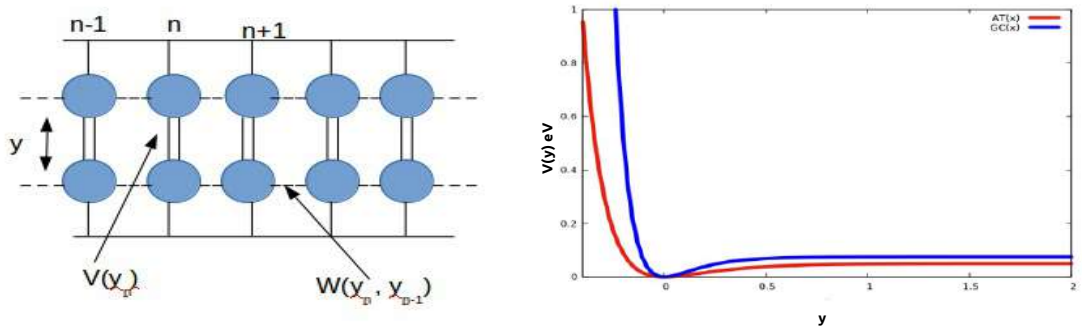


Fig. 2.3 The non-linear DNA model described by the Hamiltonian 2.2. The right panel shows the Morse potential for an homogeneous AT sequence (red line) and GC sequence (blue line). The parameters of the potential are chosen from Refs. [1, 2]

The model, schematized in figure 2.3, is described by its Hamiltonian:

$$H = \sum_n \frac{p_n^2}{2m} + W(y_n, y_{n-1}) + V(y_n), \quad \text{with } p_n = m\frac{dy_n}{dt} \tag{2.2}$$

where n is the index of a base pair and m its reduced mass. The first term, p_n^2, introduces the kinetic energy while $W(y_n, y_{n-1})$ and $V(y_n)$ correspond to potential energy, being y_n, like we mentioned above, the difference between the distance separating the two bases of the pair n and its distance of equilibrium.

The potential $W(y_n, y_{n-1})$ is an anharmonic potential which better describes the collective aspects of the transition and take into account the interaction between adjacent bases along the DNA molecule. It is known as stacking potential and accounts for the interaction between two neighboring bases at the same strand. This is the interaction that occurs between bases trying to position themselves by forming ordered structures, so that by packaging reduces its contact with the medium (an aqueous solution) surrounding the molecule. It is characteristic of the base pairs themselves, and includes different interactions like the $\pi - \pi$ orbital interaction of the base pairs plateau[19]. It should be considered like an effective potential since it include different interactions.

$$W(y_n, y_{n-1}) = \frac{1}{2}K(1 + \rho e^{-\kappa(y_n + y_{n-1})})(y_n - y_{n-1})^2 \tag{2.3}$$

the factor $(1 + \rho e^{-\kappa(y_n + y_{n-1})})$ is the one that multiplies the coupling and confers the anharmonic character. We could call the constant ρ "anharmonicity constant", since when $\rho = 0$ the coupling is harmonious and the greater ρ is, the more important is the anarmonic effect. This effect consists in changing the effective value of the coupling constant k : when $(y_n + y_{n-1}) \approx 0$ (that is, the bases $n+1$ and n are closed) the model is locally a harmonic Hamiltonian with an effective coupling constant $K' \approx K + \rho$, whereas when $(y_n + y_{n-1})$ is very large (some of the bases n or $n+1$ is open) the effective coupling constant drops to $K' \approx K$ and it is thanks to this that the sharp phase transition in long homogeneous chains can be reproduced. Indeed, the study of the model carried on Ref. [32] concludes that the phase transition of the model is first order when $\rho \neq 0$. In this way, the anharmonic factor introduces the entropic effect : when the bases are closed, the effective coupling constant is greater, which reproduces the rigidity of the double strand of DNA; when they are open, the configuration is made up of two ssDNA, much more flexible, so that the coupling constant is smaller.

$V(y_n)$ is the Morse potential, describing the repulsive interaction due to the hydrogen bonds between the two bases of a basepair on opposite strands. It takes into account the

hydrogen bonds, the interactions of the phosphates, and the surrounding solvent effects. The expression of this potential is:

$$V(y) = D(e^{-\alpha y} - 1)^2 \tag{2.4}$$

Where D is the dissociation energy of the pair, α is the width.

Figure 2.3 shows an example of this potential: for negative values the potential grows exponentially, so that prevents values too negative of y_n. This characteristic of the potential serves to model the repulsion between the backbone of sugars and phosphates of the two chains. For sufficiently large values of y_n , the potential practically reaches its asymptotic value and is flat. In this situation, the two bases have separated, the link between them has been broken and therefore there is no interaction between them, unless they come closer. The potential $V(y_n)$ is sequence dependent, so the heterogeneity of the genetic sequence is taken into account by giving different values to the parameters of this Morse potential. We have to distinguish between strong GC base pairs that are joined by three hydrogen bonds and the weak AT base pairs that are linked only by two and in which the depth and width of the potential are smaller. However, a limitation of the PBD model is that it does not distinguish between A- and T-nor between G- and C-bases. We used the parameters that in the Refs. [1, 2] were used to adjust thermodynamic properties of short heterogeneous sequences. The value of this parameters are listed in Table 2.1

The achievement of this model respect others is that PBD model describes the opening and closing of the base pairs and their evolution. Moreover, it also explains how these fluctuations occurs and their probability.

Table 2.1 Potential parameters used in the model

Model	**D(eV)**	$\alpha(\text{Å}^{-1})$	**b(eV)**	$c(\text{Å}^{-1})$	**d**	$K(\frac{eV}{\text{Å}^2})$	ρ	$\kappa(\text{Å}^{-1})$
PBD [11]	AT $\rightarrow$ 0.05000	4.2	0	\	\	0.025	2	0.35
	GC $\rightarrow$ 0.07500	6.9	0	\	\			
Adaptation [19]	AT $\rightarrow$ 0.09075	3.0	4	0.74	0.2	0.004	25	0.8
	GC $\rightarrow$ 0.09900	3.4	6	0.74	0.2			

2.2.4 Adapted PBD model

Although PBD model successfully describes the denaturation transition and the dependence of the sequence on the dynamics of the bubble formation in the pre-melting regime, i.e equilibrium properties, is qualitatively wrong for the dynamics. The average time during

which a base pair stays closed between two opening fluctuations, and the time during which it stays open after such a fluctuation are orders of magnitude too small in PBD simulations [19] in comparison to experimental values inferred from proton-deuterium exchange experiments [39]. For this reason we use an adaptation of the PBD model as proposed in ref [19]. In this work they improve the PBD model by adding an entropic barrier to the one-site Morse potential. The expression of the Morse potential is then :

$$V(y) = D(e^{-\alpha y} - 1)^2 + \theta(y)\frac{by^3}{\cosh^2[c(\alpha y - d\ln 2)]} \tag{2.5}$$

Here, b is a parameter which determines the amplitude of the barrier, c its width and d its position in units of the value of y at the inflection point of the Morse potential. $\theta(\cdot)$ is the Heaviside step-function and the original PBD model is reobtained by setting $b = 0$ in the above equation.

The existence of this entropic barrier has to be included in the potential $V(y_n)$ for a correct description of the dynamics of the open and close states. As the bases move out of the stacking, they gain entropy due to their freedom of movement once they are opened, and, in addition, they will form hydrogen bonds with the solvent. With this barrier, we can explain properly both dynamics and the life-time of the open state but also the closing transition, as this barrier takes into account the energy of the bases which flip out of the DNA stack but also the energy that the base has to overcome in order to re-close (breaking its hydrogen bonds with the solvent and reduce its entropy again) [19]. This improvement also has been tested with calorimetry and neutron scattering proving the melting transition of the A form of DNA[40].

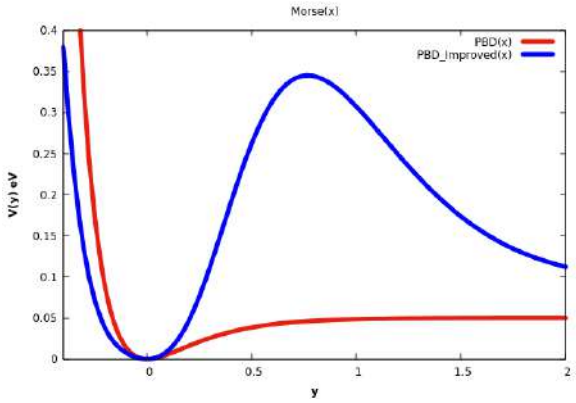


Fig. 2.4 Morse potential (red line) and the potential defined by equation 2.5 (blue line). The model using the parameters chosen for the two potentials give the same melting temperature for a poly(A) DNA.

The parameters of this adaptation of the PBD model has been chosen based on the results of theoretical calculation of stacking energies and that keep the sharp denaturation transition in DNA in agreement with experiments [19]. Figure 2.4 compares the Morse potential for the original PBD model (barrierless) and the potential of Eq. 2.5 for the same DNA homopolymer and a given stacking potential W giving the same denaturation temperature.

Both, the parameters with which the implementation PBD model are tested [19] and the parameters chosen for the original PBD model are organized on Table 2.1. The fact that $D_{GC} > D_{AT}$ reflects the reality that the G-C pairs are joined together stronger than the A-T pairs, by three hydrogen bond instead of two. In both cases studied, this reason $\frac{3}{2}$ is accurately reflected in the parameters, although it must be kept in mind that there are more factors that affect the bond between bases, since Morse is an effective potential which includes many contributions, and that the energy of these bonds that is measured in experiments does not have to exactly follow this reason. One should be very careful when trying to relate these parameters with microscopic phenomena, since it must be taken into account that in reality they represent a whole series of processes that take place at the microscopic level.

Chapter 3

DNA Hairpins

3.1 Hairpins and why is important its study

In the previous chapter we already made a first approach to the study of DNA denaturation and we defined the models that can be used for this purpose. In this chapter, we will study the denaturation of a peculiar DNA structure, called hairpin, which is a single DNA strand with a special sequence that allows the formation of this structures. We will develop a model to study this type of DNA chains and, comparing this model predictions with experimental results, we will determine what interactions and physical aspects are relevant in the dynamics of the ssDNA.

As we have said, one of the most challenging to understand DNA architectures are DNA hairpins, since they take part in many RNA configurations, playing an essential role in many biological and metabolic processes like the regulation of gene expression, DNA recombination, or regulation of the transcription by binding proteins [41–44]. We will see that the proposed model can qualitatively reproduce the phenomenology observed in the experiments with hairpins, and we will shed light on the physical origin of observed behaviors.

The hairpins are formed when we have a short single strand of DNA (not the two complementary chains that form the molecule as seen in the Fig. 1.1). This individual DNA chain has terminal regions consisting of complementary base sequences. As a result the two end-regions can self-assemble in a short DNA double helix, called the stem, while the remaining central part of the strand forms a loop [45]. A schematic example can be seen in Figure 3.1.

Hairpin structures are not static: they fluctuate between different conformations. In a simplified description, all of the conformations can be divided into two main states: the closed state, where the base pairs (bps) of the stem are paired, and the open one, where the

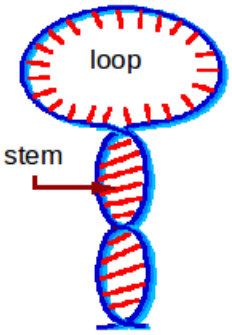


Fig. 3.1 Schematic representation of DNA haripin structure

bps of the stem are free [4]. Above a temperature T_m, called the "melting temperature", the double helix of the stem is denaturated and the molecule behaves like a standard polymer chain, which generally has its two ends far from each other [46] (See Fig. 3.2). In the closed state, the enthalpy of the system is low while in the opened state the bases are free and can adopt numerous configurations, so is a high entropy state. Thus the transition between the closed and the open state requires an energy fluctuation large enough to unzip the base pairs of the stem, which is achievable with temperature as we mentioned above.

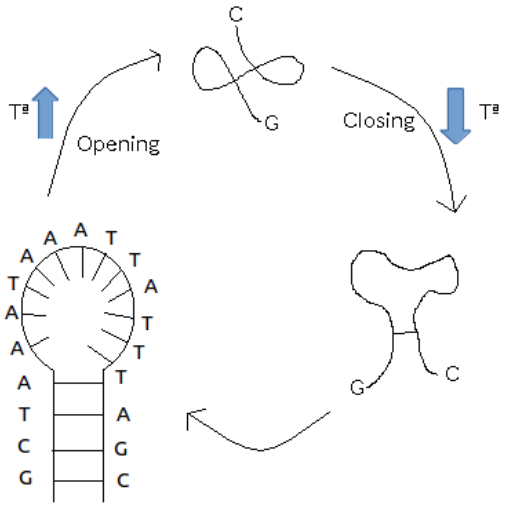


Fig. 3.2 The four bases at the two ends of the hairping are complementary to each other and the hairping flips between open and closed states with the characteristic rates

DNA dynamics and denaturation has been hardly investigated due to the experimental difficulty to measure kinetic data such as rates of opening or unwinding. Simple models of DNA, as PBD described afore, can help to analyze melting transition, but it is necessary

to validate this models by comparing with experiments. However, there has been a lack of accurate experimental data about DNA dynamics to which theoretical models can be compared. Historically DNA thermal denaturation has been investigated by ultraviolet absorption since the absorption of a sample that contains the molecule in solution is proportional to the number of bases that are open. But the denaturation rate of a hairpin is easier accessible with experimental techniques than melting of double stranded DNA, so for that, there is more experimental data available on the denaturation rate of DNA hairpins.

Nowadays, a kind of experiments that provides very accurate results with DNA hairpins have been developed. They are based on DNA probes (molecular bacons) [47, 48] which are DNA hairpins that possess a fluorophore and a quencher at their ends attached chemically. This, allow us to understand DNA hairpin self-assembly processes at the molecular scale and to detect the structure's conformational changes by an accurate monitoring of the opening and closing steps. DNA harpins are also being considered for physical applications as molecular memories (chips) [24, 49] or as engines to drive nanodevices [50, 51]

Although this kind of experiments allow to extract the kinetics of the opening/closing fluctuations, most of the studies of hairpins deal with the structure of the folded state or with the thermodynamics of the folding–unfolding transitions[45, 4]. Knowledge about the kinetics of the hairpin transitions is limited to a few isolated measurements of the characteristic opening and closing rates. However, in the last years, can be found in the literature several references of experimental measurements of the opening constant of the hairpins [4, 52, 3, 6–8, 5, 9, 53]. Our aim is to use an approach that has been used for the computation of denaturation rates of double stranded DNA [21, 54] and DNA hairpins [55], but applying the improved PBD model [19] that we defined in the previous chapter.

3.2 PBD adaptation for hairpins

DNA denaturation is a dynamical process that can be compared with experiments, but requires very long timescales that are not reachable with full-atom simulations. We have shown that an adaptation of the PBD model [19] is able to give better agreement regarding base pair lifetimes. Thus, we will investigate the dynamical properties regarding hairpin denaturation using this model and molecular dynamics simulations.

Since PBD model represents a dsDNA, to simulate hairpins it is necessary to somehow mimic the effect of the loop of the hairpin, and PBD model must be significantly extended. The full process of the closing of DNA hairpins is quite complex but two possible ways to adapt the PBD model for studying DNA hairpins have been proposed. In Ref. [45] Errami *et al* combine two different models in order to describe the hairpin. They use the PBD model

[11] to describe the stem and, the Kratky-Porod model to describe the loop[56]. The other way is the approach proposed by Hanne *et al* in Ref. [52] which is simpler and therefore with less computational efficiency. This model was already used in Ref. [55] and is based on the flexibility of the last base pair before the loop of the hairpin starts. This terminal base pair is restricted by an additional confining potential that is added to the Morse potential. The result is that its separation y cannot extend a limiting value mimicking the stretch of the loop. In Ref. [52] a square potential of $y = 50$Å is applied but in our work we use the same potential chosen by the authors in Ref. [55]. This is an exponential potential (Eq. 3.1) applied to the base pair at the end of the stem, and we use it because the exponential expression is more convenient for molecular dynamics trajectories. The expression of the exponential potential can be written as:

$$V(y_N) = \begin{cases} \dfrac{eV}{\mathring{\text{A}}^6}(y_N - \tau)^6, & \text{when } y_N > \tau \\ \\ 0, & \text{when } y_N < \tau \end{cases} \tag{3.1}$$

where τ is a hairpin cut-off. In our case the maximum separation for the terminal base pair is defined as 10Å. Definitely, the loop is not explicitly described and only the maximum distance between the terminal base pair is described. The value of the cut-off has not a big influence in the calculation of the opening rate constant of the hairpins, since it has shown that it is only slightly affected by alteration of the loop [4] while the closing rate is much more affected by changing the loop characteristics. The same has been reported with the denaturation curves which are not very sensitive to the maximum extension of the terminal base pair, and resemble closely that of double stranded DNA, *i.e.* without confining the terminal base pair[55].

3.3 Opening rate constant

Most relevant dynamical processes where DNA denaturation is involved, and there is available data that can be compared with experiments, require very long timescales. The typical time scales of the fluctuations of DNA hairpins vary from *ns* to μs so full-atom simulation is very computationally demanding. Molecular dynamics or Langevin simulations are powerful methods to simulate the movement of the molecules, estimating physical quantities, but the accessible timescale of the simulations is generally not long enough to observe the specific biological transitions such as DNA denaturation. However, thanks to the fact that the PBD model has only first-neighbor interactions, the rate constants of DNA melting transition can

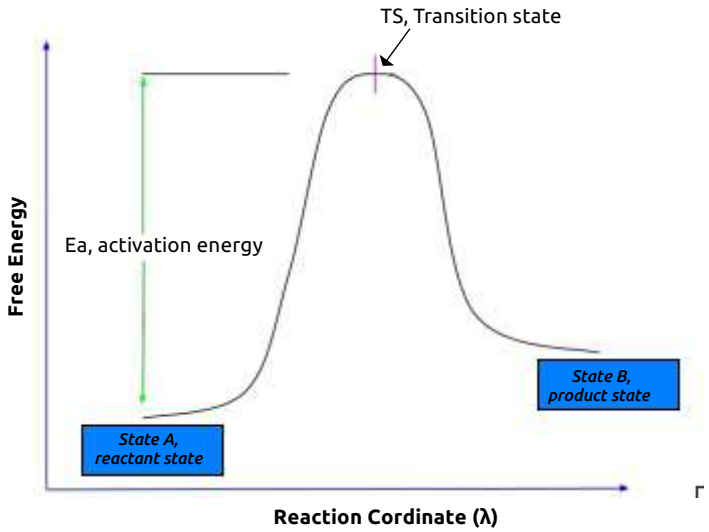


Fig. 3.3 Schematic representation of the free energy profile along the reaction coordinate of a simple reaction A⇌B. The transition state is at the maximum of the energy profile.

be computed very accurately using the approach of Refs. [21, 54, 55]. The approach is based in Transition State Theory (TST), nominated by Wigner and Eyring already in the 1930s, and a very efficient implementation of the reactive flux (RF) method which are powerful techniques to treat the PBD model. We apply this method to determine the rate constant of hairpins, that represents the fraction of closed DNA molecules that denature per units of time $(1/ns)$.

Firstly, this approach requires the computation of the free energy. To this end, first-neighbor interaction models can efficiently be computed using an iterative integration procedure[37]. This will provide the TST expression for the rate which must be corrected for correlated recrossings via a dynamical factor, the transmission coefficient. This transmission coefficient can be determined via the effective positive flux formulation [57]

3.3.1 Reactive flux method

RF method [58–60] expresses the overall reaction rate (in our case the opening of the DNA hairpin) as an equilibrium of density probability of being at the surface of the barrier. This barrier is one which the reactants must overcome in order to the reaction to be carried out and to separate two stable states (See Fig.3.3). The method starts from the expression of the TST, but corrects it for correlated recrossings via the transmission coefficient [21]. Then, the rate constant can be written as the following product:

$$k = k^{TST}\, x\, R \tag{3.2}$$

Here, k^{TST} is the transition theory expression for the rate constant that the RF method provides as an exact expression that is independent of the choice of reaction coordinate (RC). RC was chosen as $\lambda \equiv min[y_i]$ like in previous works [21, 54] so it is defined by $\lambda(y_i) = y_0$ with y_0 the equilibrium distance ($y_i > y_0$ implies that bp i is open).

With the condition that the system is at the reactant side of the barrier, the reaction rate expressed as a probability is:

$$k = P(\lambda = y_0 \mid \lambda \leq y_0)\, x\, R \tag{3.3}$$

where P is the probability density and R is a dynamical factor called the unnormalized transmission coefficient and always has a value between 0 and 1. This factor corrects for fast correlated recrossing avoiding the problem of overestimating the rate constant and it is calculated by releasing dynamical trajectories forward and backward from the top of the free energy barrier. Thus, finding the rate constant now consist of two basic operations, the calculation of the free energy and the calculation of the dynamical transmission coefficient. There are several methods to calculate both types of quantities and in Ref. [54] authors propose how to do it very efficiently by exploiting specific peculiarities of the PBD model.

Free energy calculation using direct numerical integration method (DNIM)

Due to first neighbor character of the PBD model the free energy of the probability density can be computed very efficiently with an iterative numerical integration scheme. This method, if can be used, is more efficient than others. It is based on the conversion of multi-dimensional integrals of properties as probability density, into products of low-dimensional integrals. This low-dimensional integrals can be solved by iteration [37]. We can make use of the fact that if the integrant is of a factorisable form, we can solve the integral

$$Z = \int dy^N a^{(N)}(y_N, y_{N-1}) \ldots a^{(3)}(y_3, y_2) a^{(2)}(y_2, y_1) \tag{3.4}$$

using the following iterative scheme:

$$z^{(2)}(y_2) = \int dy_1 a^{(2)}(y_2, y_1)$$

$$z^{(3)}(y_3) = \int dy_2 a^{(3)}(y_3, y_2) z^{(2)}(y_2)$$

$$z^{(N)}(y_N) = \int dy_{N-1} a^{(N)}(y_N, y_{N-1}) z^{(N-1)}(y_{N-1})$$

$$Z = \int dy_N z^{(N)}(y_N) \tag{3.5}$$

This is a huge improvement and we can optimize even more the calculation by introducing convenient cutoffs for the integration boundaries.

The equilibrium density is related to the free energy by $F(\lambda) = -k_B T ln P(\lambda)$ being k_B the Boltzmann constant and T the temperature in kelvin. Therefore, the probability density $P(\lambda = y_0 \mid \lambda \leq y_0)$ can be expressed as:

$$P = \frac{\sum_i \int dy_1 \ldots dy_N \delta(y_i - y_0) \prod_{j \neq i} \theta(y_j - y_0) e^{-\beta U(y_i)}}{\int dy_1 \ldots dy_N (1 - \prod_k \theta(y_k - y_0)) e^{-\beta U(y_i)}} \tag{3.6}$$

where $U(y_i)$ is the Hamiltonian for the adapted PBD model and thus can be solved by DNIM because all integrals are in the factorisable form.

Now we have to define the integration conditions. We select $y_0 = 2.5\text{Å}$ as threshold value, in the plateau of the Morse potential where the base pair is totally open. The integration step chosen is $d_y = 0.05\text{Å}$ which is small enough to minimize systematic error without increasing too much the computational cost. The integration boundaries used that make the integration stops are

$$\mid y_i - y_{i-1} \mid < d = \sqrt{2 \mid ln\varepsilon \mid / \beta K}$$

$$-1/\alpha_{AT} ln[\sqrt{\mid ln\varepsilon \mid / \beta D_{AT}} + 1] < y_i < y_0 + \sqrt{N} d \tag{3.7}$$

setting tolerance ε to 10^{-40} which makes all arrangement giving a contribution of $e^{-\beta V(y_i)}$ smaller than ε will be disregarded.

Calculation of the transmission coefficient by the effective positive flux (EPF) formalism

Like we mentioned above, the transmission coefficient is a dynamical factor that corrects for fast correlated recrossings [61] and that was introduced by Keck [62]. Although several methods exist to calculate the transmission coefficient, we choose the effective positive flux (EPF) that has shown to be the most effective [63]. The EPF formalism calculates the transmission coefficient R by counting, for each crossing, which one goes in the right

direction and which one is "effective". In this context, an effective crossing is that which divides the surface λ^* for the first time, and that leaves state A to go to state B without coming back to A (See Fig. 3.4).

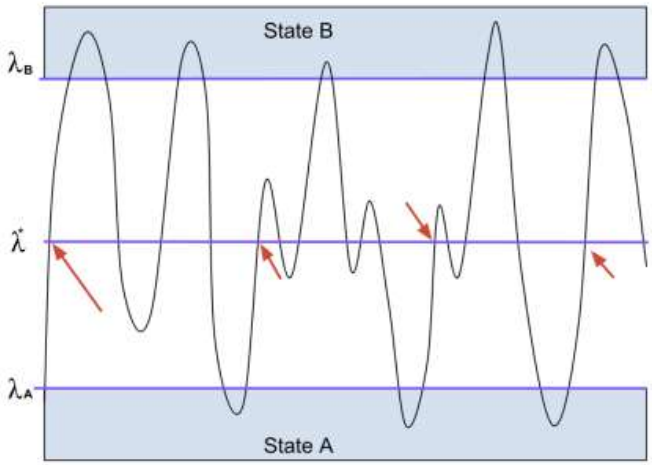


Fig. 3.4 Counting of the EFP formalism for the calculation of the transmission coefficient. Arrows mark an effective crossing which have assigned value 1 to h^b_{0,y_0}. Not pointed out crossings have $h^b_{0,y_0} = 0$

The dynamical factor transsition coefficient is expressed as:

$$\langle \dot{\lambda}\,\theta(\dot{\lambda})h^b_{0,y_0}\rangle_{\lambda=y_0} \tag{3.8}$$

h^b_{0,y_0} is a binary function that is 0 unless it is a effective positive crossing, which then is equal to 1. The transmission coefficient is obtained by running dynamical trajectories (in our case 10^6, using a time step of 1 fs and bp masses of 300 amu.) with random velocity according to the Maxwell-Boltzmann distribution. One should kept in mind that for our PBD model, we assume that the system is in B state when is at the plateau of the Morse potential, *i.e* $\lambda^* = \lambda_B = y_0$. Finally, it is important to stress that a proper choice of the RC is critical for the efficiency of the RF method. There are alternatives to the RF method as Transition Path Sampling (TSP) based method, but for our system RF method is extremely fast.

Besides, we select Langevin dynamics to perform the simulations which is a very commonly used technique. Langevin describes the effect of a solvent assuming that the molecules feel a friction due to the viscosity of this solvent. We use a friction coefficient of $\gamma = 50ps^{-1}$ which is considered an appropriate value for the friction of water [21, 64]. Temperature is also controlled with this dynamics.

Molecular Dynamics simulations

Summarizing, we will examine the dynamical properties of DNA denaturation by calculating the opening constant via performing the molecular dynamics simulations explained above. The simulations have been carried out by a fortan code written by Titus van Erp and implemented by Oda Dahlen and myself during an internship at Norwegian University of Science and Technology (NTNU) of Trondheim, Norway. For each simulation, the program gives the free energy along the reaction coordinate, the transmission coefficient and thus, the rate constant.

For DNA hairpins, we compare our results using the recent adaptation of the PBD model [19] with experimentally obtained values found in the literature. In addition to the model itself we examine some other variations with different parameters to explore the effect on the behavior of denaturation rates as function of temperature and chain length. Following the same theory, instead of hairpins with which the model is not very accurate due to the potential that mimics the loop, we want to apply our model to dsDNA chains and study its bubble length size and distribution.

3.4 Results and Discussion

DNA Hairpins

Since we consider the problem of the fluctuation dynamics of hairpins is still unknown, we want to make a theoretical analysis and compare it with experimental available results, in order to determine the basic mechanisms controlling the properties of DNA hairpins. The main problem is to understand the mechanism of hairpins transition and our aim in this study is to introduce a simple approach that allows us to identify the basic phenomena involved in the statistical and dynamical properties of DNA hairpins. Regarding theoretical studies, other models have been used to explain this phenomenology [46, 45] and also PBD model has been previously used with qualitative results [52, 55].

In Ref. [52] Hanne *et col.* present a single-molecule measurements of the opening rate of DNA hairpins under mechanical tension and compare it with the results obtained from a reduced-degrees-of freedom statistical mechanics model (PBD). In this case, constants rates of short hairpins were computed in a qualitative manner using an artificial Monte Carlo type of dynamics. We use the RF approach described above, that was already used by Dahlen and van Erp to report quantitative computations of hairpin denaturation rates [55].

The results of the calculated hairpin denaturation rate constant for different sequences for which experimental values are available, obtained using the RF methodology explained above and the implementation of the PBD model (Eqs. 2.3-2.5) are detailed below.

Table 3.1 shows the sequences of the hairpins investigated and temperatures of the different data points in Fig. 3.5 together with their experimental reference.

Table 3.1 Sequences and temperatures of thermal unfolding hairpins

Label	**Stem**	**Loop**	**Temp.**(°C)	**Ref.**
S_1	AAGGG	T_{21}	25	[3]
S_2	GGGAA	T_4;A_{30}	23	[4, 5]
S_3	AAGGAA	T_4	54	[6]
S_4	GGAAAA	T_4	23	[7]
S_5	GGGAAAA	T_8	36	[8]
S_6	GAGAAGA	A_{40}	25	[9]
S_7	GAGAAGAGA	A_{40}	25	[9]

In Fig. 3.5 only original PBD model [11] with parameters from Campa and Giansanti [1] and the model with the additional barrier of Ref.[19] are shown. It can be observed that with our implementation of the PBD model we can almost obtain the same order of magnitude than the experimental hairpins.

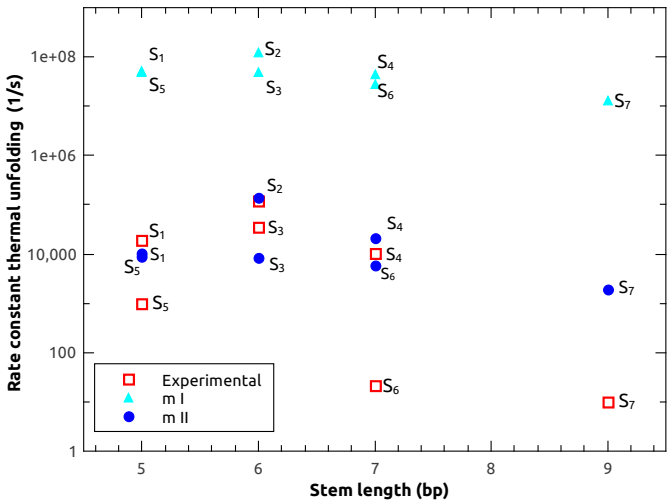


Fig. 3.5 Opening rates of hairpins in Refs. [3–9]. The picture shows the results of the original barrierless PBD model and PBD model with barrier compared with the experimental results for sequences listed in Table 3.1 Statistical errors of the simulations are of the size of the maker.

The agreement with experimental results is therefore much better than with the original PBD model. PBD model is three till five orders of magnitude too high while the adapted model is either on top of the results for short sequences or maximally two orders bigger for the longer sequences. It is important to realize that also the experimental data have some differences for sequences that have the same stem length and same GC content. One should kept in mind that there are a lot of factors that can affect the dynamics of the hairpins including differences in loop length or sequence, variations in the temperatures under which the experiments were performed or the measurement technique itself.

Table 3.2 Potential parameters used in the model

Model	**D(eV)**	$\alpha(\text{Å}^{-1})$	**b(eV)**	$c(\text{Å}^{-1})$	**d**	$K(\frac{eV}{\text{Å}^2})$	ρ	$\kappa(\text{Å}^{-1})$
mI	AT $\rightarrow$ 0.05000	4.2	0	\	\	0.025	2	0.35
	GC $\rightarrow$ 0.07500	6.9	0	\	\			
mII	AT $\rightarrow$ 0.09075	3.0	4	0.74	0.2	0.004	25	0.8
	GC $\rightarrow$ 0.09900	3.4	6	0.74	0.2			
mIII	AT $\rightarrow$ 0.09075	3.0	6	0.74	0.2	0.004	25	0.8
	GC $\rightarrow$ 0.09900	3.4	9	0.74	0.2			
mIV	AT $\rightarrow$ 0.09075	3.0	8	0.74	0.2	0.004	25	0.8
	GC $\rightarrow$ 0.09900	3.4	12	0.74	0.2			
mV	AT $\rightarrow$ 0.1255	3.0	4	0.74	0.2	0.004	25	0.8
	GC $\rightarrow$ 0.1455	3.4	6	0.74	0.2			
mVI	AT $\rightarrow$ 0.1031	3.0	4	0.8	0.3	0.004	25	0.8
	GC $\rightarrow$ 0.11785	3.4	6	0.8	0.3			
mVI-a	AT $\rightarrow$ 0.1031	3.0	4	0.8	0.3	0.001	25	0.8
	GC $\rightarrow$ 0.11785	3.4	6	0.8	0.3			
mVI-b	AT $\rightarrow$ 0.1031	3.0	4	0.8	0.3	0.004	2	0.8
	GC $\rightarrow$ 0.11785	3.4	6	0.8	0.3			
mVII	AT $\rightarrow$ 0.1031	3.0	4	0.74	0.2	0.004	25	0.8
	GC $\rightarrow$ 0.11785	3.4	6	0.74	0.2			
mVII-a	AT $\rightarrow$ 0.1031	3.0	4	0.74	0.2	0.001	25	0.8
	GC $\rightarrow$ 0.11785	3.4	6	0.74	0.2			
mVII-b	AT $\rightarrow$ 0.1031	3.0	4	0.74	0.2	0.004	2	0.8
	GC $\rightarrow$ 0.11785	3.4	6	0.74	0.2			
mVIII	AT $\rightarrow$ 0.1155	3.0	4	0.8	0.3	0.004	25	0.8
	GC $\rightarrow$ 0.1355	3.4	6	0.8	0.3			

Besides the latter models, we also use some variations of the model applying addaptations of the parameter sets in order to study the trend in the denaturation rate when the shape of the

added barrier and the Morse potential is varied. Table 3.2 shows each parameter set which defines a case model, notated as mI-mVIII.

As we defined previously in Table 2.1 mI and mII are respectively PBD original model [11] and the adaptation of Ref. [19] with a barrier in the Morse potential. In models mIII and mIV we increase the barrier height by a factor of 1.5 and 2 respectively, in order to study how the shape of the barrier affects the denaturation rate. Finally, models V-VIII have a variation in the depth of the Morse potential (or what is the same, we changed the height of the plateau in different proportions) which leads to a change in the probabilities between the open and closed state for individual base pairs. In addition, models mVI and mVII present in turn a variation in the strength (a) and in the stiffness (b) changing parameters K and ρ of Eq. 2.3 respectively [65]. All the parameter sets checked keep the shape of the potential with the entropic barrier, except mI that is original barrierless PBD, like shows Fig.3.6.

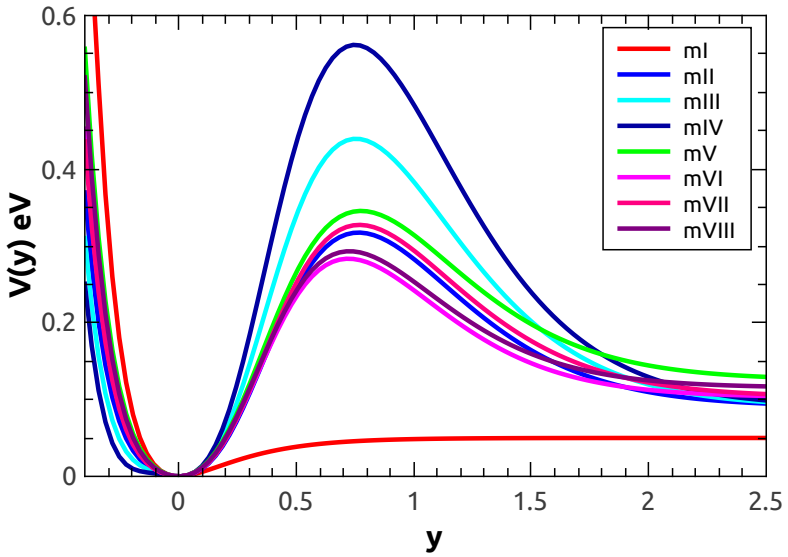


Fig. 3.6 Morse potential for the models listed in Table 3.2. The potentials are drawed for an homogenoeus AT sequence.

We have seen that barrierless PBD model [11] only obtain qualitative results in the determination of opening rate constants of hairpins denatured thermically, and the improved PBD model with an entropic barrier [19] explains better theoretically the dynamics of short molecules. Now, we repeat the simulations to obtain denaturation rates of hairpins but using the models which involve some parameter fitting. Fig. 3.7 depicts the result of the calculated rate constants with different models listed in Table 3.2 for the hairpins in Table 3.1 and comparing it with the experimental values.

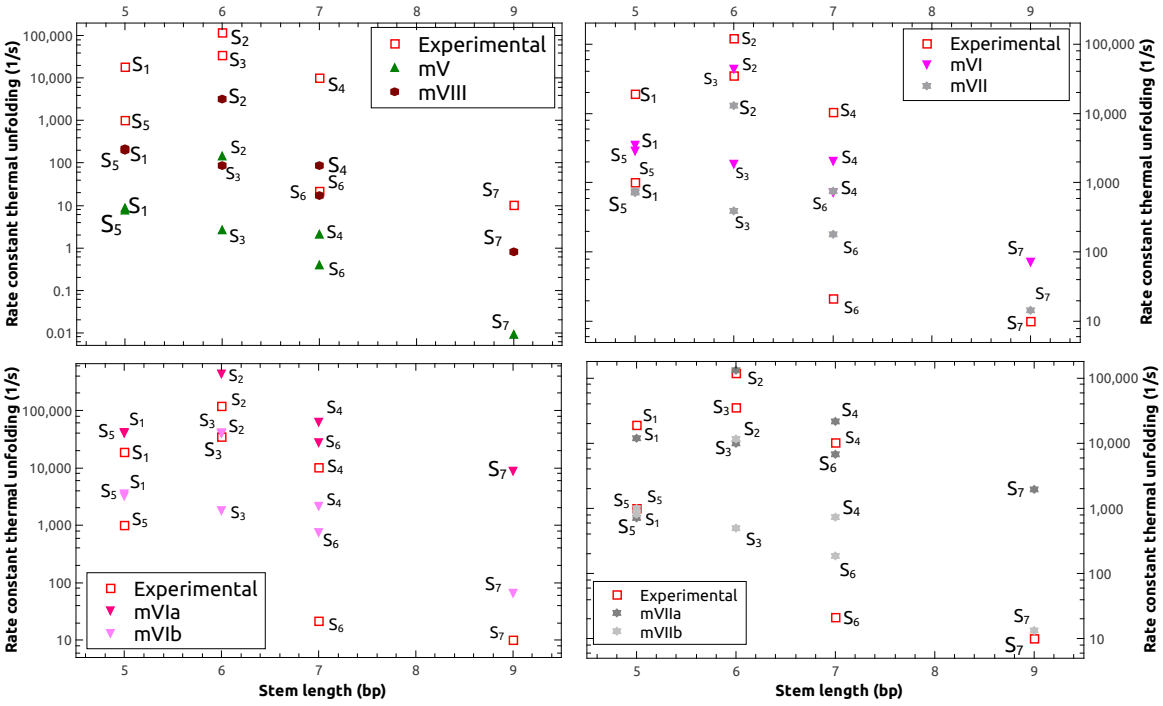


Fig. 3.7 Opening rates of hairpins in Table 3.1. The picture shows the results of the calculated opening rate using all parameter sets tested.

As it can be observed, we obtain good agreement between some experimental results and some of the models, but neither of them it is capable of reproducing all experimental values. Some models are nearly as good as mII but none of them is better, so the only thing clear is that the addition of an entropic barrier in the Morse potential improves the description of the rate openings for DNA hairpins. Other conclusion is that whatever the model used, the agreement with experimental results is always better for molecules with short sequence stem. While the stem rise up, the effect of the loop of the hairpin plays a more important role and we can see that with big stems the model has worse behaviour. This is because of a cooperative effect that is not included in the model despite of the barrier.

We examine the denaturation rate as function of temperature looking for experimental results that were obtained in this way. In Ref. [4] Bonnet *et al.* investigated the denaturation rate constant as a function of temperature for two hairpins with identical five base-pair stem GGGAA but with different loops, one having a loop with 21-bps of pure T bases and the other one 21-bsp of pure A bases. Here, one should take into account that for our model, this two hairpins have identical description. It does not distinguish between A and T bases (for our model the AT bp has the same physics as the TA bp) but also because the loop itself has not explicitly description as we explained above. Once again, firstly we use models mI and

mII of table 3.1 which do not involve any parameter fitting and then we calculate the opening rates of this hairpins with the models mIII-mVIII. Once again, we obtained good quantitative results like we can observe at Fig. 3.8. The experiments were performed at temperatures ranging from 280 to 315 K.

The results of Fig. 3.8-a) indicate that the original PBD model is far from the experimental results by five orders of magnitude at low temperature and three orders at high temperatures. However, with the model with the barrier in the whole range of temperatures, we are at the same order of magnitude or one order over than the experimental rates. Focusing in Fig. 3.8-b) and c), it can be observed the same experimental results again together with the models mIII-mVIII. Picture b) shows that putting the barrier high (mIII and mIV), the rate constants of the hairpins are smaller. The barrier takes into account entropic effects but also a enthalpic contribution and other effects. Rise up the barrier does more difficult a bp to open but also the opened base pairs have more difficulty to reclose. This contradictory effect shows that the barrier in the potential is necessary but not enough since reclosing rate is also affected. It is necesary also playing with the high of the plateau in the potential. We can see in Fig. 3.8-b) and c) that with the barrier and rising the plateau (mV-mVIII, see Table 3.2 for the parameters), we can obtain the same order of magnitude of the experimental results in all the range investigated (mVI). The difference is that some models give rate constant that are lower than the experimental results, while others results are higher. The slope of the curves of the opening rates as function of the inverse temperature on the logarithmic scale, is nearly the same. As with the earlier experiment, this shows that the model with barrier proposed in Ref. [19] considerably improves the dynamics of DNA and, especially for short DNA sequences, reproduces the experimental denaturation constants reasonably well. In addition, it can be concluded that playing with the high of the plateau of the potential is really important in the physical and dynamic description.

On the other hand, if we now compare the denaturation constants versus stem length of the hairpins, we see a different behavior. We examine the hairpins in Ref. [53], where individual DNA hairpins were unzipped and re-folded at constant force by using an ultrastable, high-precision optical force clamp. Although it is not the same case as we denatured hairpins by increasing temperature, the estimation range of the opening rates may be representative. Hairpins contain an 50% GC base pairs, a 4T loop sequence, and stems ranging in length from 6 to 30 base pairs. The stem sequences are given in Table 3.3

All experiments were performed at 23 degrees Celsius (296 Kelvin), and rate constants were obtained. We applied the PBD model and the implementation model of Ref. [19] with all parameters sets in order to obtain rate constants for these hairpins, and compare it with

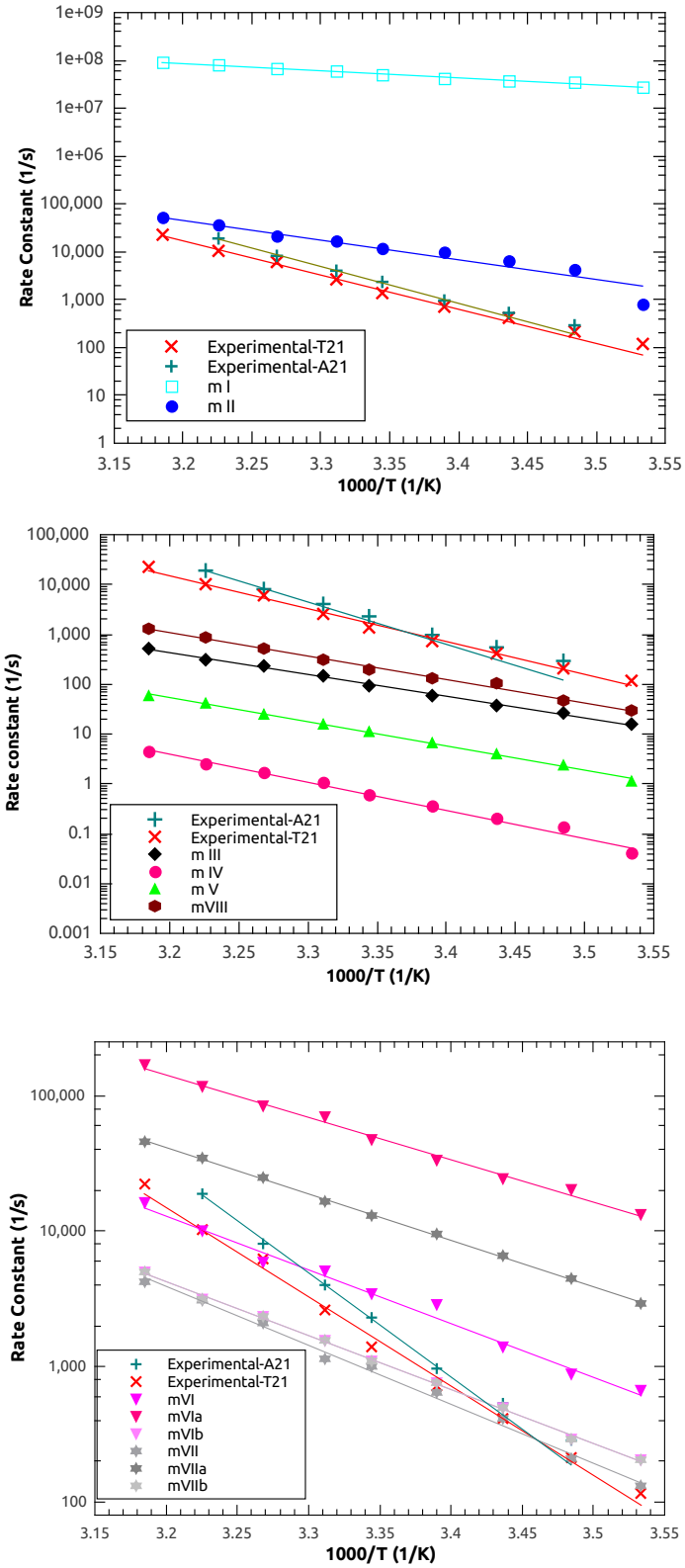


Fig. 3.8 Opening constant of hairpins with stem GGGAA and loop of T21 and A21 as a function of the inverse temperature and comparison with our model with different sets of parameters. a) Experimental data compared with theoretical results from the original PBD model (mI) and the adapted PBD model (mII). b) and c) Comparison of the same experimental data with the theoretical results of the adapted PBD model with parameter variations (models mIII-mVIII of table 3.2)

Table 3.3 Sequences and experimental rate constant of the investigated hairpins in optical tweezer experiment

Sequence	**Stem length(bp)**	**Rate Cons.** $(\frac{1}{S})$
GAGGAA	6	11.023
GAGAGGAA	8	1.491
GAGAGAGGAA	10	0.016
GAGAGGAGGAAGGAA	15	$6.14E^{-6}$
GAGAGAAGGAGAGGAAGGAA	20	$2.54E^{-13}$
GAGAGAAGGAAGAGGGAGGAAGGAA	25	$8.53E^{-17}$
GAGAGAAGGAAGAGAAGAGGGAGGAAGGAA	30	$9.60E^{-24}$

experimental results. Fig. 3.9 shows the dependency of denaturation constant on stem length for seven experimental hairpins.

As it can be observed in Fig. 3.9, both the theoretical model and the experiments show an exponential decrease of the denaturations rates as a function of stem length. The lines in Fig. 3.9 are exponential fits through the linear part in the log-plot. Original PBD model has a much weaker decay as function of chain length which reflect a deviation between the model and the experimental results about beyond 7 orders of magnitude for the shortest chain (6 base pairs) and almost 30 orders of magnitude for longest lengths (30 base pairs). Adding the barrier (Ref. [19]), allows to estimate accuratelier the order of magnitude of the opening rate constants calculated. The model with the barrier with original parameters set (mII) is around 5 orders of magnitude lower than the original PBD model for all sequences, i.e. closer to the experimental results, but it also fails to reproduce the same slope in the ln k versus chain length plot.

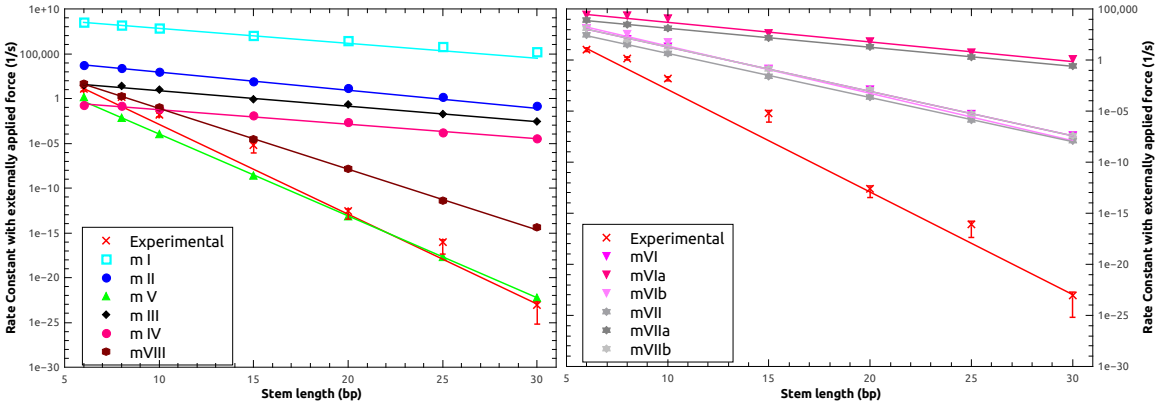


Fig. 3.9 Calculated denaturation constant by thermal simulation with models mI-mVIII of Table 3.2 and compared with rate constant measured by unloaded optical tweezer experiments.

Changing the shape of the barrier, i.e making it higher, do not improve the results as we can see observing models mIII and mIV in which we increase the barrier height by a factor 1.5 and 2, respectively. Also it can be shown that the models which present a stiffness variation (mVIb and mVIIb) improve a little the slope, being closer to experimental results.

In order to get qualitatively correct behavior of the opening constant as a function of stem length, it is necessary to change the high of the plateau of the potential, or what is the same, change the depth of the Morse potential. Model mVIII and even better model mV can quantitatively reproduce the optical tweezer experiments at all lengths. Having said that, to change the depth of the Morse potential makes the denaturation curves somehow unreal, since they acquire a form non-adjusted to realistic data. Fig. 3.10 represents melting curves for homogeneous homopolymers showing the typical phase transition [11]. The melting temperature at which the transition occurs depends on different factors (such as the solvent, ion concentration...) but experimentally this occurs around 70 °C for pure GC sequences [66]. However, the model mV, despite giving the best agreements with the experimental denaturing rates of hairpins, gives melting transitions for AT and GC chains quite far for the typically values, that are even beyond the boiling temperature of water.

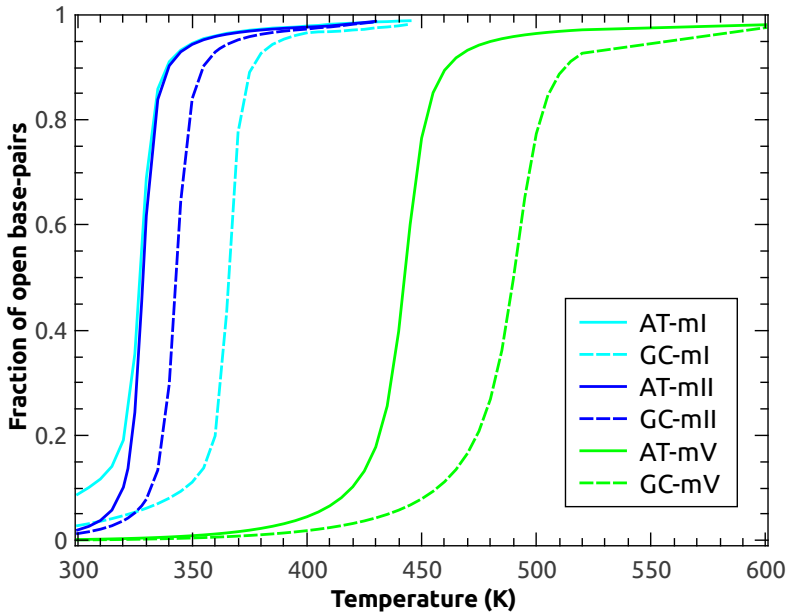


Fig. 3.10 Phase transition represented as a fraction of open base pairs as function of temperature for long (400 base pairs) homopolymers calculated using the approach of Ref. [10] with models mI, mII and mV of Table 3.2

.

As it can be observed in Fig. 3.10 model mII reproduce the melting temperature of long homopolymers (87 °C for model mII and 92 °C with original PBD model, mI). We showed that with this recent adaptation [19] which add an additional barrier to the Morse potential, the agreement of the denaturation rates for DNA hairpins in comparison with experimental results, can be improved. With the experiments that we compared, we have shown that for short sequences this agreement is relatively good but with longer sequences it is more difficult to explain. The magnitude of the exponential decay of the rate as function of chain length does not correspond to those measured by optical tweezer experiments [53].

Therefore, it is a problem to get at the same time correct dynamics and thermodynamics using a single mesoscopic model. One issue that might repair both rate trends and denaturation curves is the missing rotational degrees of freedom. It could be done by replacing the flat horizontal Morse plateau with a function having negative slope to describe entropic effects not considered yet. This results suggests that in order for the PBD model to describe dynamics correctly, fundamental changes have to be made to the model. Other possible alternative is to use a more complex model and fit it to experimental data since full atomistic simulations have a large computational expense. Improving mesoscopic models for DNA is a continuous challenge for researches in the field but it is also necessary more accurate experimental data.

Double-stranded DNA and bubble length distribution

Hairpins are a very particular architecture of DNA involved in many biological processes, but as we explained in the first chapter, DNA which contain genetic information, is a double-stranded helix. Following the same theory as until now, instead of hairpins with which the model is not very accurate due to the several factors that we explained during this chapter, we want to apply our model to dsDNA chains. Firstly, we use again the RF approach described above and do MD simulations with PBD implemented model [19] in order to obtain denaturation rate for a sequence containing a bubble in the middle. Then, we also calculate the distribution of bubble lengths in double-stranded DNA for segments of varying guanine-cytosine (GC) content and present its analytical description. This is based on recent experiments [67–69] in which the role of intermediate states in denaturation is studied.

The sequence for which we calculate the opening rate, has 26bp (GGGGAAAAAAAAAA AAAAAAAGGGG) and we start with a big bubble. We will keep doing the bubble smaller by changing one adenine in the middle of the sequence for one guanine each row, until the whole sequence is made of guanines. In this way, at the beginning we depart from a sequence with a big bubble, after that we have intermediate states with two small bubbles

(we are closing the bubble from the middle to the ends) and finally the sequence without many bubbles. The simulations were performed at temperatures ranging from 310 to 380 K. Fig. 3.11 shows the rate constant calculated with the adaptation of the PBD model [19] (mII) as a function of increasing temperature.

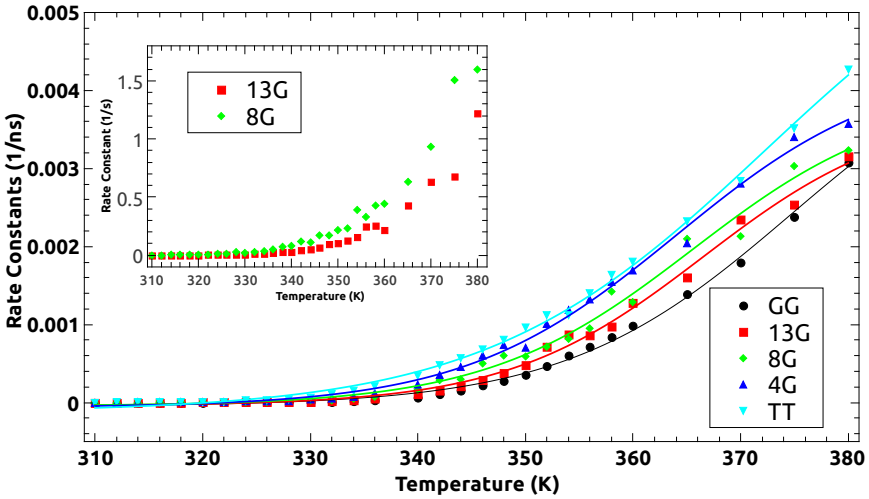


Fig. 3.11 Rate Constant of a dsDNA sequence with a AA bubble in the middle as a function of increasing temperature and fitted with a Boltzman function. Inset shows the result of the same calculation for the seuqence with 8G and 13G with the PBD original set of parameters (mI,[11].)

.

We can observe that when we have the big bubble (TT) or we don't have any (GG) the sequence denatures in a single step and it is due to the fact that the curves present only one slope. But when we have intermediate states, like when we have 8G or 13G changed in the bubble, and with enough temperature for the bubble to be nucleated, the rate constant is not in a single step, if-not there is two different regimes and two different slopes in the curves between a plateau. This behavior was also found in calorimetric experiments with the same sequence and the same result can be observed in Fig. 3.12 which shows the Boltzmann coefficients for the fit curves.

In this figure (3.12), it can see that the sequence with all A in the bubble or the one with all G, has a high coefficient corresponding with the big slope of the curve, and the coefficients of the curves between those two, are smaller due to the fact that the transition is not in a single step. In the inset of Fig. 3.11 we can see that this behavior is not observed in the same sequence but with the original PBD model (without the entropic barrier in the Morse on-site potential [11]), so the implementation model (mII) [19], allowed to represent more accurately

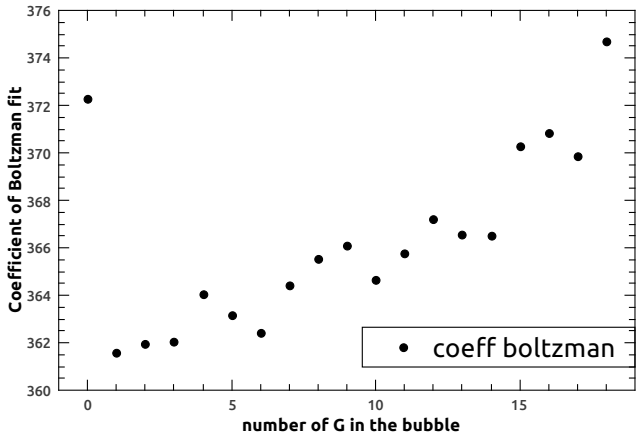


Fig. 3.12 Boltzmann coeficients for the fit curves in Fig. 3.11

.

what occurs biologically. In conclusion, this adaptation also improves the description of the rate openings for dsDNA chains.

In order to explain the bubble length distribution in dsDNA, we first should describe the methodology used and some concepts. The experiment in which this procedure is based, allows to differentiate molecules that are completely denatured from those that are only partially open. The experiment consists in the following: after having a solution of the DNA molecule of interest, in thermal equilibrium at the desired temperature, the sample is cooled suddenly and in a really abrupt way. Due to the fast decrease in temperature, the individual chains of ssDNA present in the sample from completely denatured dsDNA molecules, do not have time to recombine with complementary individual strands, and they close on themselves forming hairpins. The dsDNA molecules that were only partially open, close again. Here we can define p as the proportion of molecules that are completely denatured at the studied temperature (open molecules) and f the fraction of open base pairs. We introduce characteristic functions that indicate if a certain base-pair is open or closed [10]:

$$\theta_k(y_k) \equiv \theta(y_k - \xi), \qquad \overline{\theta_k}(y_k) \equiv \theta(\xi - y_k) \tag{3.9}$$

Here, ξ is the definition of the opening threshold and it indicates that a configuration is double strand when $y_k < \xi$ and to the contrary is completely denaturated whenever $y_k > \xi$ for all k. Regarding $\theta(\cdot)$ is the Heaviside step-function that we already shown in the adapted PBD model and that is equal to 1 if the base-pair is open and is zero otherwise. Once defined

θ_k we can give mathematical expressions to the proportion of open molecules and base pairs :

$$f = \frac{1}{N}\sum_{k=1}^{N}\langle\theta_k\rangle$$
$$p = \langle\prod_{k=1}^{N}\theta_k\rangle \tag{3.10}$$

Based on these magnitudes, p and f, we can know properties of the intermediate states (bubbles)which happen in the denaturation. Starting from this, we can easily calculate $\langle l\rangle$ what we define as the fraction of open base-pairs provided that the molecules is in the double stranded state, and sometimes is called the average fractional bubble length.

$$l = \frac{1}{N}\sum_{k=1}^{N}\langle\theta_k\rangle_\mu, \qquad l = \frac{f-p}{1-p} \tag{3.11}$$

Another interesting observation is the fraction of bases that are participating in intermediate states, σ. Its value is obtained simply by subtracting the fraction of open base pairs to the fraction corresponding to denatured molecules, in such a way that:

$$\sigma = f - p \tag{3.12}$$

Finally, we define the bubble probability matrix $P_{bub}(k,m)$ as the probability to have a bubble of size m centered at base-pair k when the molecule is in dsDNA state, and this matrix contains all the information of the bubble statistics in a DNA sequence. Hence:

$$P_{bub}(k,m) \equiv \langle\theta_k^{[m]}\rangle_\mu \tag{3.13}$$

Both, $\langle\theta_k\rangle_\mu$ and $\langle\theta_k^{[m]}\rangle_\mu$ that appear in Eqs. 3.11-3.13 can be expressed using partition function integrals and therefore are factorisable integrals which can be solved with the direct numerical integration method explained above.

$$\langle\theta_k\rangle_\mu = \frac{Z_{\theta_k} - Z_\pi}{Z - Z_\pi}$$
$$\langle\theta_k^{[m]}\rangle_\mu = \frac{Z_{\theta_k^{[m]}}}{Z - Z_\pi} \tag{3.14}$$

We have studied the same sequences that in Ref. [70] consisting in 200bp segments of varying guanine-cytosine (GC) content, using MD simulations of Hamiltonian 2.2 with the adaptation of the Morse potential (Eq. 2.5 with parameters of Ref. [19]). We discuss the distribution of bubble lengths at $T = 310K$ and apply Periodic Boundary Conditions (PBC) which means that the chain is 3 times longer but we compute only the middle part. A problem with the Hamiltonian of Eq. 2.2 is that, if all the base pairs are separated enough to escape the well of the Morse potential, there is nothing to favor the two chains joining again, and the y_n variables diverge [71] preventing simulations to be long enough to obtain a reasonable statistic. Also, as we did before, we introduce the agreement to consider that a pair is open when $y_n > 1.5$Å.

Figure 3.13 shows the bubble length distribution per base pair, $P(l)$, obtained from our MD simulations with the bubble statistic procedure explained above. This bubble length distribution per base pair is defined as

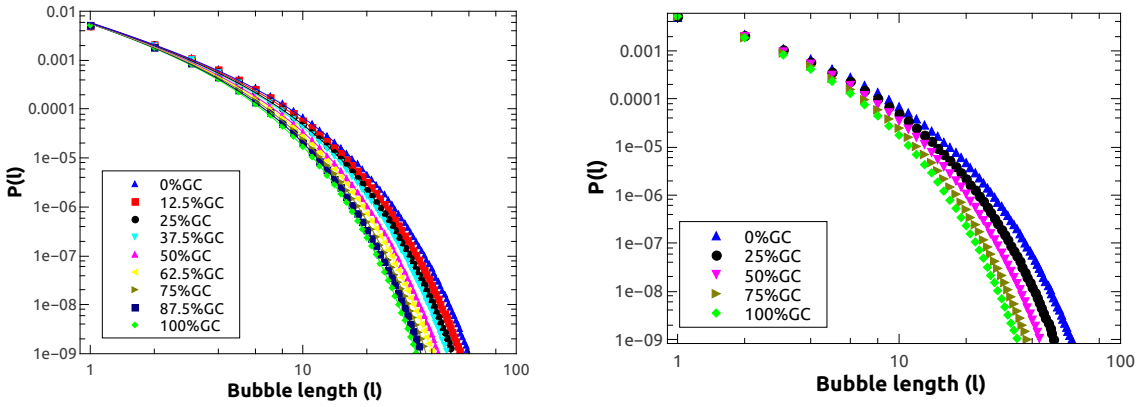
$$P(l) \to P(l) = Lim_{L\to\infty}\frac{< N(l) >}{L} \tag{3.15}$$

averaged number of bubbles of length l base pairs on a sequence of total size L. The distribution $P(l)$ should be interpreted in the following terms: for a specific DNA sequence of total length L base pairs, the quantity $P(l)\cdot l$ gives the average number of occurrences of a bubble of length l base pairs, in thermal equilibrium. It has been seen that the curves can be fitted with a nonexponential law:

$$P(l) = W\frac{e^{\frac{-l}{\xi}}}{l^c} \text{ with } l \geq 1 \tag{3.16}$$

The results of the obtained bubble length distribution per base pair, $P(l)$, using Ref. [19] parameters (Adapted PBD model, mII) are shown in Fig. 3.13

Using the improved PBD model with the barrier [19], it obtains different orders of magnitude than Ref. [70] but it was expected. The difference is one order of magnitude lower for small bubbles and two orders of magnitude smaller for big bubbles. Also, we can observe that all sequences studied have slower decay with the PBD barrier model. Bubble tails with



(a) All seuqences with fit

(b) Few sequences without fit

Fig. 3.13 Results obtained with PBD model with entropic barrier, a) all sequences studied (random GC content from 0% till 100%) and fitted using Eq.2.2. b) The same picture with a lower number of sequences studied and without fit.

model with the barrier, show more exponential tendency (contrary to with Original PBD model study in Ref. [70], that are more linear) due to the difficulty of reclosing. Therefore they are more time opened so the life time of the bubbles is longer with the barrier.

Next step was to quantify the change of the distribution with the GC content. The dependence of the parameters of equation 3.16 on the GC fraction is shown. In the next pictures, it is plot together the parameters of the fits obtained with PBD with barrier (mII) [19] and the ones obtained in Ref. [70].

Data extracted from the Ref. [70], that corresponds to Monte Carlo simulations with original PBD model (mI)[11], give a tendency and a fit in two distinct regimes with linear functions (open light blue squares). This two distinct regimes are above 37.5% GC and below that. But observing the parameters obtained fitting Fig. 3.13 corresponding with PBD model with barrier (mII), give a tendency completely different (dark blue circles).

For Ref. [70], preexponential coefficient W decrease monotonically with the GC percentage. The decay of W signifies that the higher the GC content the more difficult it is to excite large openings in the double strand; therefore in AT-rich sequences bubbles have a higher statistical weight. But in our case, with the model with the barrier, W increases with the GC content. This means that it is easier to excite large openings in the dsDNA, contrary to what Ref. [70] suggested. Bubbles excitation by temperature is characteristic of the model with the barrier. Finally, in the model with the barrier $P(0)$ is really similar for each %GC so it seems that the obtained tendency of preexponential coeficient W it is logical.

Exponential coefficient c also decrease monotonically with the GC percentage in Ref. [70]. In this case, with the model with the barrier, we obtain lower parameters but whit the same

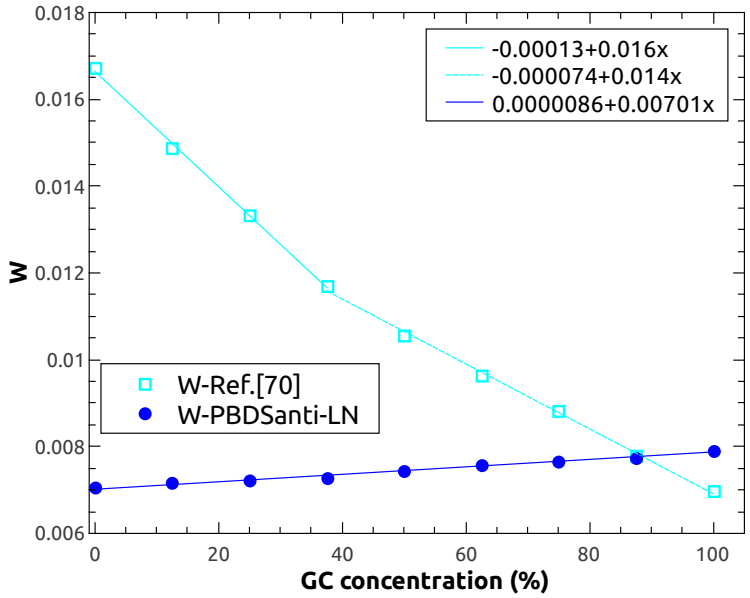


Fig. 3.14 preexponential coefficient W from the fits using Eq. 3.16 .

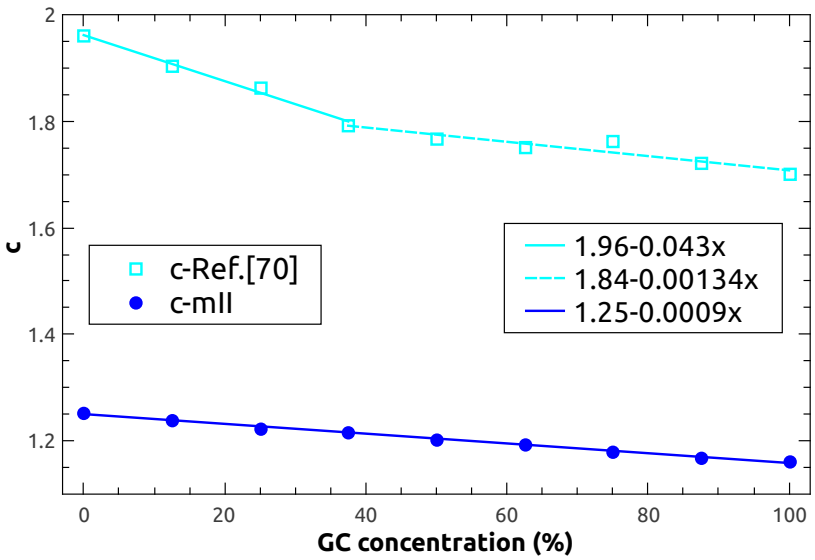
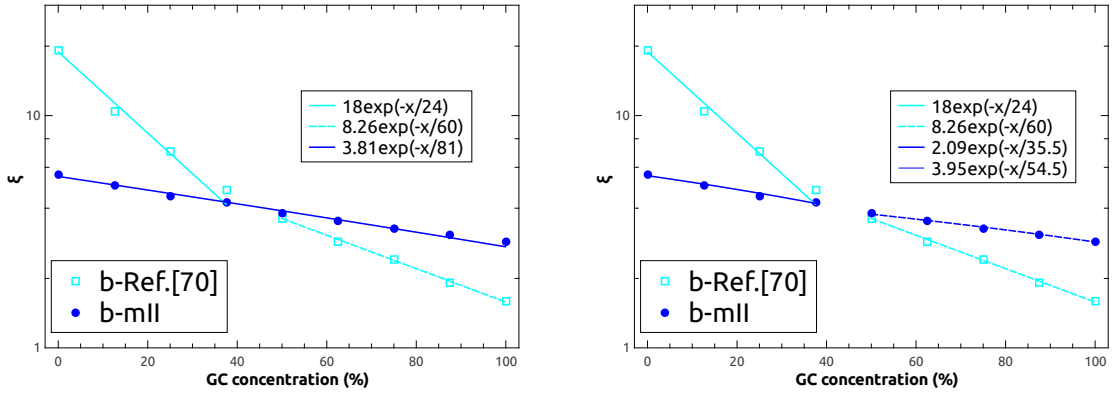


Fig. 3.15 exponent c from the fits using Eq. 2.2.

decreasing tendency. The exponent coefficient c has again a behavior consisting in two distinct regimes on the GC fraction dependency: above and below a GC content of about 40%. With PBD barrier model, c decreases, but we cannot see different regimens. Exponent coefficient c, also has the meaning of the order of the phase transition. For Ref. [70], the c coefficient has a value of $3/2 < c < 2$, meaning a second order melting transition while in the model with the barrier, $1 < c < 3/2$ indicates a smoother transition.



(a) Single regimen

(b) Two distinct regimes

Fig. 3.16 decay length ξ parameter from the fits using Eq. 2.2 . a) Parameter ξ fitted in a single regimen and b) parameter ξ fitted in 2 distinct regimes below and above 37.5%GC

Now, we can see that decay length ξ is closer to Ref. [70] results (PBD Original model), but some differences exist between them. The decay length ξ is smaller for GC-rich sequences, in which the distribution decays faster. However, we can see that although with PBD barrier model, decay length ξ also decrease monotonically with GC percentage, the transition is softer.

Previous studies [72] have shown that the nucleation of bubbles depends strongly on the sequence: the weaker AT base pairs have to go over the potential barrier imposed by their GC neighbors in order to break the bonds. Our resultss suggest that for lower concentrations of GC, the formation of bubbles is not a GC-dominated process, and large AT regions can form bubbles freely, hence dominating the bubble formation process.

Next step is to calculate the average bubble length L_B, which is given by the total number of base pairs in bubble states divided by the total number of bubbles

$$L_B = \frac{\sum_l lP(l)}{\sum_{l\geq 1} P(l)} \tag{3.17}$$

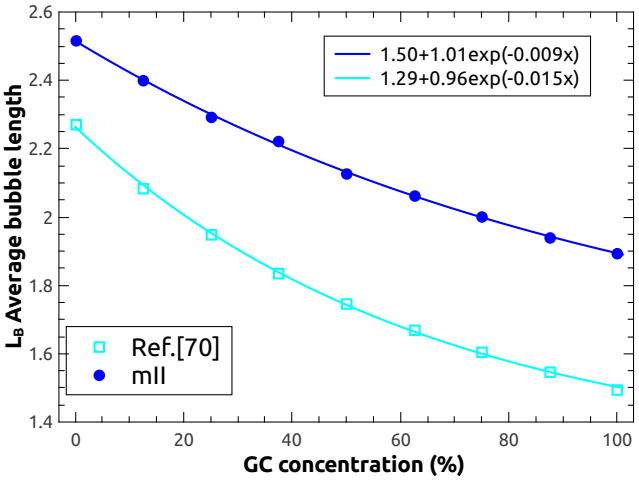


Fig. 3.17 Dependence of the average bubble length L_B, Eq. 3.17, on the GC content of the sequences.

With the model with the barrier, we obtain the same exponential decay but with higher avarage bubble. L_B depends strongly on the GC content, showing an exponential decay. This stresses the importance the sequence has on the typical size of denaturation bubbles.

Until now, we have studied sequences with different content of GC but the distribution of the GC% in the sequence was random, so now, we repeat the experiment putting the AT% in the middle of the sequence, being the bubble in the middle. The probability of bubble length distribution per base pairs when the bubble is in the middle of the sequence is shown in Fig. 3.18

We can observe that the tendency is a priori the same that Fig 3.13 , with the GC percentage distributed randomly, but there are some differences. The ranges of 0-12.5% of GC, 25-37.5% of GC and 50-62.5% of GC have very similar bubble length distribution per base pair while higher GC contents have more different probability. Also, the parameters of the fits change, as we can see in Fig. 3.19. Hence, in the next pictures we plot together the parameters obtained with the adapted PBD model for sequences with random GC content (random in the legends) and with the bubble in the middle of the sequence (bm indicator) in order to compare both between them, besides with the results of Ref. [70].

As it can be observed in Fig. 3.19, with adapted PBD model (mII) in the sequences with the bubble in the middle we can see a different behavior than the shown with the random distribution. Now we can observe 2 distinct regimens like Ref. [70] suggested but in a different way. Above 40%GC W increase with GC content, but below this percentage the tendency of W is more linear.

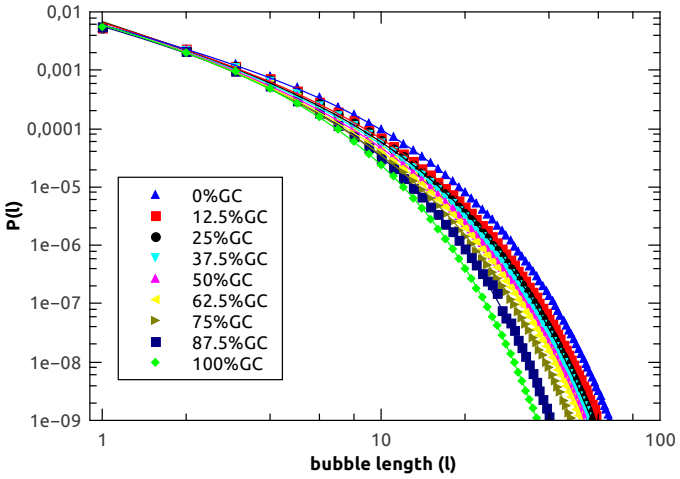


Fig. 3.18 Results obtained with PBD model with the barrier and with sequences with bubble in the middle.

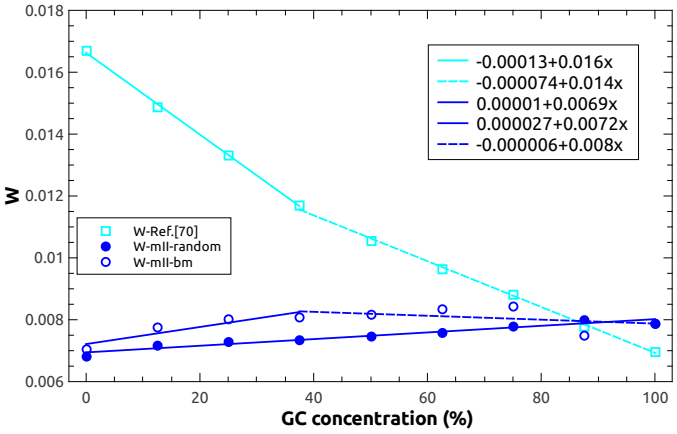


Fig. 3.19 preexponential coefficient W for sequences with %GC distributed randomly and sequences with bubble in the middel obtained with PBD model with barrier.

Now, Fig 3.20 shows completely different tendency in the exponent c between sequences having GC content randomly distributed and sequences with bubble in the middle. In the model with the barrier, c in sequences randomly distributed decreases with GC content while c in sequences with bubble in the middle increases until a high %GC around 75% where drops abruptly.

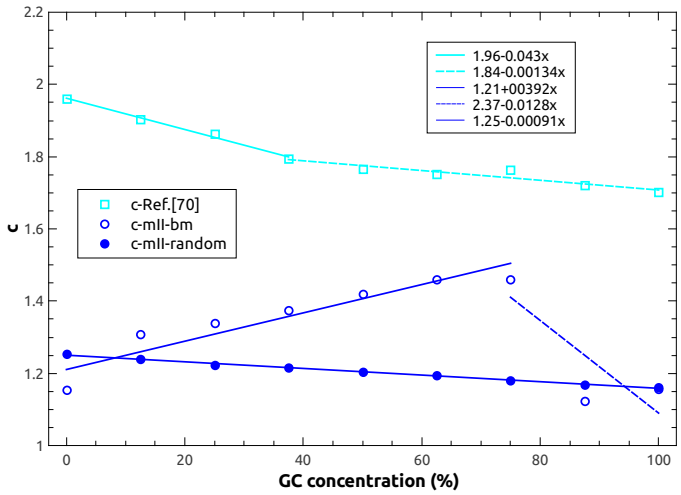


Fig. 3.20 Exponent c for sequences with %GC distributed randomly and sequences with bubble in the middel obtained with PBD model with barrier.

In the decay length parameter, we also observe differences between sequences with GC percentage distributed randomly and sequences with the bubble in the middle. While in the sequences with GC content distributed randomly, decay length ξ decreases monotonically with GC percentage but in a softer way, for sequences with the bubble in the middle, decay length ξ decreases slowly or keeps more or less constant until a percentage of GC around 60% where decrease faster.

The bigger difference between the sequence with GC content distributed randomly and the sequence with the bubble in the middle is that while the first presents an exponential decay, the second one follows a polynomial distribution.

Summarizing, we have shown that the PBD model implemented with an entropic barrier into the Morse on site Potential, reproduce reasonably well experimental results of DNA denaturation for short sequences, but differs a little in the description of the bubble distribution and intermediate states. It is due to the presence of the barrier, which makes more difficult reclosing base pairs and once bubbles excitation is produced, this is more lasting in time. Even so, this shows that the ingredients of PBD model, the interaction between the bases through the hydrogen bonds and the role of the entropy difference between the closed phase of the molecule (dsDNA) and the open (ssDNA), due in large part to the interaction of

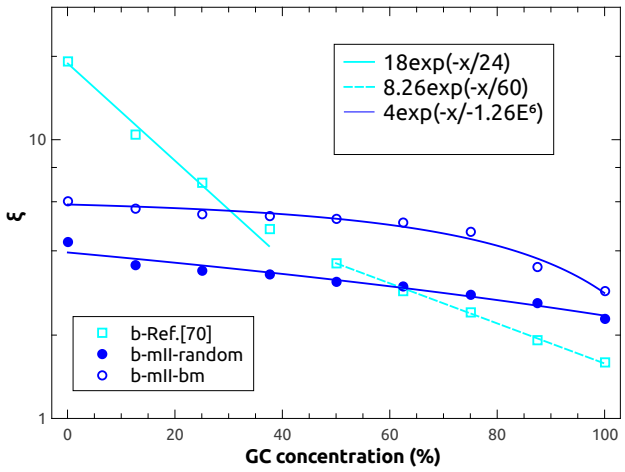


Fig. 3.21 Decay length ξ for sequences with %GC distributed randomly and sequences with bubble in the middel obtained with PBD model with barrier. .

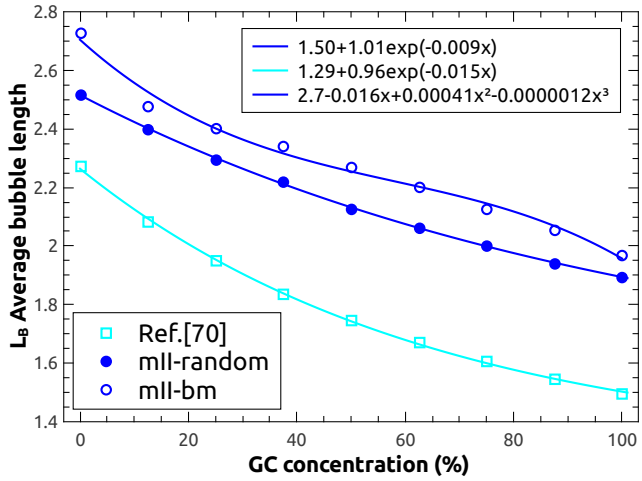


Fig. 3.22 Dependence of the average bubble length L_B, Eq. 3.17, on the GC content of the sequences

stacking in the dsDNA, are sufficient to explain not only the denaturation, but even details such as the formation of bubbles.

Chapter 4

Neutron Scattering for the study of DNA denaturation

4.1 Neutron Scattering

Neutron scattering is a powerful technique for the study of structural and dynamic properties of different materials (condensed phases) at the atomic and molecular level. Neutrons have a dual nature, wave-particle, which makes them suitable for diffraction following the de Broglie relation

$$\lambda = \frac{h}{m_n v} \tag{4.1}$$

Neutrons were found by Chadwick in 1932 and since then the creation of neutron sources for the research of this particle and its applications has been promoted. In order to produce neutrons, a nuclear reactor (fission) or spallation source (conversion of heavy nuclei) with suitable monochromators which select the λ by Bragg diffraction, is usually used to select the desired neutron wavelength. The Institute Laue-Langevin (ILL) in Grenoble, France, where we have carried out the experiments made in the present doctoral thesis, began in 1972 to operate as one of the most powerful sources in the world [17].

Neutrons are uncharged particles that can interact directly with atomic nuclei. They have a magnetic moment (spin $\frac{1}{2}$) that can interact with the magnetic moment of the particles of interest of study. In addition, neutrons have energy similar to the energy of elementary excitations and its wavelength is comparable with the distance between atoms in a crystalline lattice (atomic inter-spaced).

Some of its advantages over other scattering techniques are the location of hydrogens and light atoms, magnetism, contrast matching (for biology specially) and magnetism which is extremelly difficult to access with X rays.

In neutron scattering experiments, part of the neutrons inciding on the sample are transmitted, some are absorbed, and the rest are scattered. Figure 4.1 shows the schematic process of a neutron scattering.

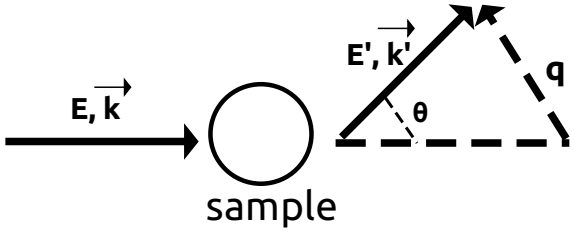


Fig. 4.1 General scheme of a neutron scattering experiment

Here, E and E' are the energies of the incident and scattered neutron respectively, k and k' their moments and q is the so-called scattering vector.

Neutron scattering is a non-destructive technique, that is, it does not cause damage to the material studied. But also has an inconvenient which is the large amount of sample necessary to carry out the experiments due to the low probability of interaction of neutrons with matter.

Neutron diffractions can be classified into elastic and inelastic and each of them is composed of coherent and incoherent diffraction depending on the type of sample (nuclei) that we have.

Elastic Scattering

During the elastic scattering the nucleus is endowed with a total kinetic energy superior to that which it had at the beginning. In an elastic event there is not effective change of the energy of the neutron ($E = E'$) or the magnitude of the momentum($k = k'$). Only the direction of the momentum changes. This is so in the center of mass frame. Elastic scattering occurs when the nuclei of the atoms irradiated with neutrons are light; and normally the applied neutrons will have low-energy. We did not used energy analyser in the experiment but it is common to assume that most of the scattering with thermal neutrons is elastic.

When slow neutrons with wavelengths comparable to the distance between the nuclei of a substance are scattered by a block of matter, two kinds of phenomena are presented: incoherent scattering and coherent or Bragg scattering.

The coherent signal appears as the result of the correlation of the different nuclei at different times and of the correlations of a nucleus at different times. It is an interference and a collective phenomenon that allows the determination of the probability af an atom being in a position with respect to the rest of the sample, in which are included: Bragg Peaks, phonons and spin wave.

The incoherent dispersion has its origin in the correlations between the positions of the same difractor element at different times. It happens when the nuclei are at varying distances and in complete disorder such as in a gas or in a totally amorphous liquid; then the effects of the waves scattered by each nucleus are totally independent. It is a non interference phenomena that allows to reflect single-particle scattering as atomic diffusion or vibrational density states.

Therefore, a neutron scattering experiment will be composed by coherent and incoherent parts.

The Bragg dispersion is a coherent and elastic scattering process. It occurs in crystalline samples whose nuclei are regularly arranged in a reticular geometry, and is characterized because neutrons are emitted under certain privileged angles that correspond to the angles given by the Bragg formula (Eq. 4.2):

$$n\lambda = 2d\cos\theta \tag{4.2}$$

where d is the distance between crystallographic planes, n is an integer, θ is the angle formed by the incident neutron with the scattered one and $\lambda = 2\pi/k$ is the wavelength associated with neutrons.

A Bragg peak contains more information than simply its position. The analysis of the shape and width can determine its correlation function and characteristic length.

One of the experiments in which we could be present at ILL as part of this thesis, is the use of neutron scattering to measure the temperature dependence of a strong Bragg peak of a B-DNA fiber sample immersed in different solvents.

Inelastic Scattering

In this process the neutron also loses energy. However, the lost energy will be used not only to supply kinetic energy to the nucleus, but also to excite it ($k \neq k'$). Consequently, the nuclus will remain in a state superior to fundamental. In general, this energy is remitted in the form of electromagnetic radiation by means of one or several photons. In the inelastic dispersion, a change in the wave vector occurs. These magnitudes are related to the law of inelastic scattering $S(q,w)$.

Inelastic neutron scattering provides information about dynamics and also gives energy spectra as a function of Q with high resolution which allows to understand spatial vibrations of the sample. Depending on coherent and incoherent scattering, two important phenomenons could be studied, phonon dispersion curve and phonon density of states respectively.

Neutron scattering is used in many scientific fields. Regarding biophysics, our team has been carrying out neutrons experiments at the ILL on DNA for a number of years. The possibility of having DNA in fibres (oriented DNA molecules) have opened new opportunities for experimental studies. The next section of the chapter provides a description of how to obtain this kind of highly oriented fiber DNA samples, which has been used in some neutron scattering experiments, besides to be the samples used in the experiments of studying the melting transition of DNA by Raman spectroscopy carried out in this thesis. In addition, we also will explain the preparation of a DNA solution sample containing the Widom sequence, used to study the flexibility of short DNA chains by small angle neutron scattering (SANS).

4.2 Sample preparation

4.2.1 Highly oriented DNA fibers

Oriented samples are really necessary for structural studies. Long DNA molecules can form DNA fibers, where the molecules adopt a close-packed arrangement. Afterwards, DNA fibers can be reoriented to form highly organized samples, such as DNA films. Within DNA films, fibers are disposed adjacent and parallel with the molecules organized in a periodic lattice leading to crystalline regions in which DNA is oriented along the molecular axis. It is also possible to find periodic positioning of the molecules in the perpendicular direction of the fiber axis. In the present work the analyzed samples consisting in DNA films are made by Wet Spinning technique [73–75].

The wet spinning method consists in precipitating DNA in a salt solution with high ethanol concentration where DNA is not soluble. The reason for this insolubility is the

difference between the low dielectric constant of ethanol compared with the higher one of pure water. The solubility also depends on the concentration of salts present in the solution. With high concentrations of ethanol and in presence of positive ions, the dielectric constant of the medium decreases, leading the ions to attach to the DNA molecules. Because of that, negative charge of DNA surface decreases favoring intermolecular aggregation.

The samples were made from an aqueous solution of DNA sodium salt from salmon testes from Sigma-Aldrich. This DNA has a GC content around 40% and a counter-ion was used: sodium or lithium. For the Raman experiments it was used mainly Na-DNA fibers while for neutron scattering experiments both types of fibers (Na-DNA and Li-DNA) were utilized. For Raman samples, the solution used in order to get DNA fibers was a tris buffer solution with $1.58g/L$ of DNA, $0.15M$ NaCl, $0.003M$ Na-citrate and $5 \cdot 10^{-5}M$ Na-EDTA into a high concentrated EtOH solution (75% in $0.02M$ NaCl) [76, 77]. EDTA is a chelating agent that can form complexes with divalent cations such as Mg^{2+}, which function as cofactors of many enzymes such as exonucleases, thereby reducing the activity of such enzymes and acting like inhibitors. For Li-DNA fibers a similar solution consisting of $1.58g/L$ of DNA, $0.3M$ LiCl, and complementary salts as $0.2g/L$ of sodium azide and $0.02g/L$ of EDTA, was used.

The final DNA film is obtained after cleaning, drying and re-humidificating processes in a 75% relative humidity in order to set the water content, stabilize the structure of the film and made the film removal easier.

Wet Spinning apparatus

Figure 4.2 depicts the schematic structure of the wet spinning apparatus used to elaborate samples of DNA fibers used in the different experiment this thesis is related with. The wet spinning apparatus is a homemade and almost unique in the world device which was donated to the ILL in 2004 by H. Grimm.

Before starting, it must be taken into account that the formation of bubbles during the process should be avoided since they could lead to ruin the preparation of the sample. To do this, all liquids are degassed for 4 hours in a vacuum pump. Once ready, the glass container and the degasser are filled with a mixture of water/ethanol and with the help of a vacuum pump also the glass column, creating a gradient from the base to its upper part. In the case of Na-DNA fibers, the mixture is 75% ethanol with 0.02M NaCl while for Li-DNA fibers it is 83% ethanol with 0.4M LiCl. DNA solution is extruded through the spinneret at the top of the glass column where the ethylic solution is located, at a rate of 33 ml/h. The spinneret is a piece with many small orifices that mimics the effect of a microscopic shower forming fibers with molecules pre-oriented that precipitate along the column in contact with concentrated ethanol and finally converge in the lower area of the column, in a V-shape wire. The V-shaped

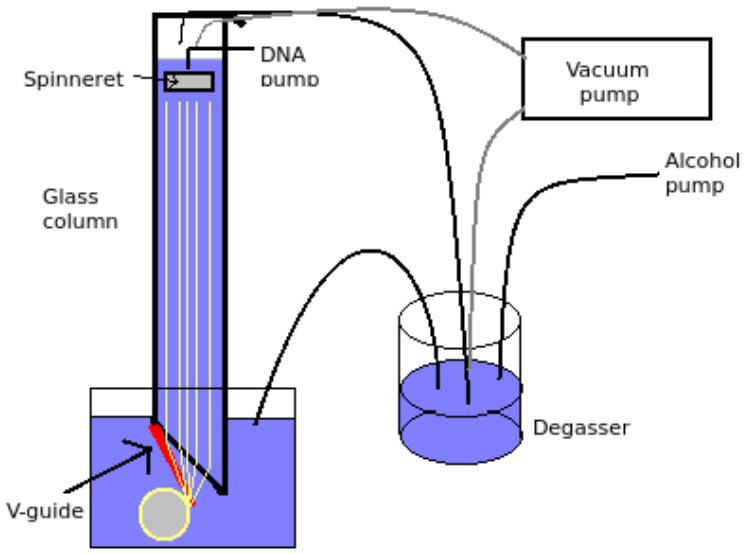


Fig. 4.2 Scheme of the wet spinning apparatus.

wire guides the resulting fiber onto a spinning teflon-coated cylinder, where the fiber is rolled up. The cylinder also undergoes a translation in a direction perpendicular to the winding fiber, which will help us to control its width. During this process, which can last from 2 to 5 hours depending on the initial volume of DNA, there is a recycling system. It is done by injecting a more concentrated solution of ethanol and it is necessary to avoid the dilution of the solution when water is added in the upper part, and to prevent the precipitation of the DNA for being interrupted.

In the whole process it is really important the position of the V-guide respect to the spinneret and the cylinder, which surface needs to be smooth in order to avoid fiber stacking. The orientation of the final spun fibers is due to different contributions according to Ziabicki *et al.* [78, 79]: certain orientation is attained while aggregated DNA is flowing down in the column and also during the drying process of the film.

Once the films are made, the final DNA ordered sample is obtained after cleaning, drying and re-humidificating processes like we said above. Fig. 4.3 shows a picture of the final result of DNA films.

Uses of DNA fibers

DNA film samples made as described above have been used in two different types of experiments. Several experiments using neutron scattering have been carried out at the ILL facilities, for the study of the melting dynamics of DNA fibers under a humidified atmosphere (9-13-625 and 9-13-678). Making use of the instruments D16 and D19, the evolution of

Fig. 4.3 Final result of the DNA films obtained with wet spinning apparatus.

the Bragg peaks of DNA fibers has been followed to determine how the correlation length changes through the transition and to correlate these changes with structural changes in the DNA denaturation. In December of the last year, a set of Li-DNA samples submerged in a water/ethanol mixture was studied varying the percentage from 20 to 80% of ethanol. This samples gave us some issues and so far only one sample was preliminary analyzed. They looks interesting, but the analysis is still a work in progress. Also the influence of spatial confinement on the structural correlation during the melting transition has been investigated in Na-DNA fibers immersed in solutions of poly(ethylene glycol) (PEG).

Samples for neutron experiments should consist in several films with a total mass of at least 0.6g of DNA in order to obtain a good signal to noise ratio. DNA films are introduced in an aluminum cassette with their axis coaligned. DNA is totally submerged in around 0.8mL of the solution of interest previously degassed.

DNA films are also used for RAMAN experiments where the melting transition of the DNA in fiber has been studied paying attention to the resonance information of the Raman frequencies of the DNA base pairs. We will discuss these experiments in more detail in chapter 5.

4.2.2 DNA short chains (Widom sequence)

DNA bending is an important factor in the biological functions of the molecule, including how it packs and its interaction with proteins and membranes. The stiffness or flexibility of DNA, quantified through the definition of the *persistence length*, is a property that dictates much of the conformational forms and behavior of the molecule. The persistence length is a parameter that may be used to model how DNA interacts with itself, and thus it is a vital and useful parameter to understand many biological properties since to perform their biological

functions,macromolecules must adopt a multiplicity of conformations. The persistence length varies in the presence of denaturation bubbles an trough the melting transition of DNA.

As we have mentioned in the first part of this chapter, our working team has been the first one in reporting structural information through the melting transition of B-DNA fibers by mean of neutron diffraction, measuring the spatial correlation along the molecule, and achieving in this regard a success where other techniques failed [20]. We have reported the necessary information to determine the size of the regions that stay in the double-helix conformation as the temperature increases until reaching the melting temperature. Now, we propose to use SANS (small-angle neutron scattering) to follow the shape of short-length DNA with predefined sequences as a function of temperature, thus correlating the persistence length to the formation of bubbles. The ultimate aim of this experiment is once again to investigate the statistical dynamics of the DNA molecule.

The double helix structure of DNA is rather rigid, with a persistence length of 500Å[80, 81], which is sufficiently flexible to allow the molecule to pack in structures like the genome in eukaryotic cells while still supporting the propagation of phonons [82, 83]. However, the persistence length may not be homogeneous along a DNA molecule. As recently shown by members of our team [38], the persistence length may change locally due to the sequence and to external effects such as temperature. Changes in the persistence length may have important biological consequences since balancing and controlling different conformational states is fundamental to biological processes. In a clear example, some sequences of DNA are known to have histone positioning effects when they are wrapped into chromatin. A local variation in persistence length that was better adapted to the curvature required could explain the histone preference for those DNA sequences; offering then a logical explanation to: histone positioning and DNA wrapping. DNA fitted in a nucleus of $10\mu m$ diameter is a extreme packing performed firstly 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. Figure 4.4 presents the diagram of a nucleosome.

The isolation of sequences that have a strong affinity to bind to the histones [84] raises the possibility of studying the structure of DNA wrapped around histones [85] and to get some insight on the mechanics behind DNA sequence-dependent properties of the nucleosome [86]. However, these studies, which are able to give detailed information of the structure of the DNA once it interacts with the histones, do not give information of whether some intrinsic mechanical and geometrical properties of DNA are also playing a role in the mechanisms that are responsible for the strong positioning effect.

One of this sequences with well known affinity to bind around the histones is the so called Widom-601 sequence discovered by Lowary and Widom [84], which has 145 bps with a total

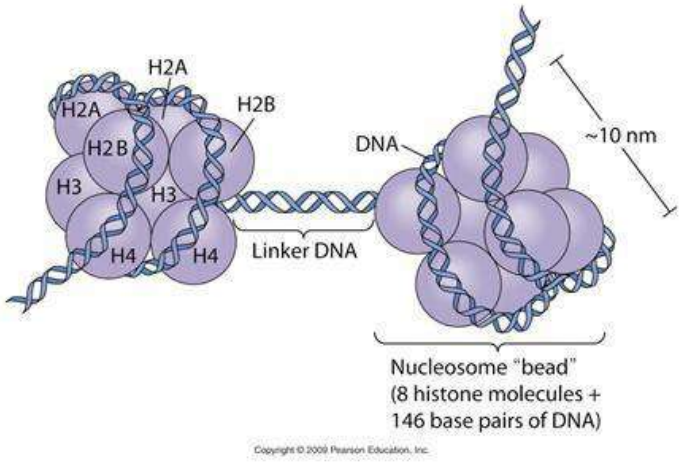


Fig. 4.4 Diagram of a nucleosome. Reproduced from the webpage http://bioloudla.blogspot.com.es/

length of around 485Å. The artificial synthesis of this sequence has created the possibility of studying the structure and parameters of the binding in the nucleosome core particle (NCP). The main aim of this part of the present thesis is to develop a specific synthesis methodology (and beginning its legal register) for the design of short-chain synthetic DNA, with controlled sequence. This technique will allow us to study the stiffness of DNA by neutron scattering in order to understand some highly important but still open projects, like the local flexibility of DNA [87]. Once the sample is made, it will be used in small angle scattering of neutrons (SANS) and x-rays (SAXS) experiments in order to study the properties of this Widom-601 sequence.

The synthesized DNA sequence was free in solution, in a salt/deuterated water, and thus was free from perturbations inherent in previous similar studies such as binding to fluorophores or gold nanoparticles. During the development of this thesis, a synthetic DNA fragment with the relevant biological characteristics of Widom sequence but higher length has been built, being in this case the most optimal configuration a DNA fragment formed by two widom sequences putted together. The considerable amount of artificial DNA needed for the neutrons and x-ray experiments was synthesized following the next strategy of genetic engineering. The starting raw material was the plasmid pGEM-3z/601 purchased from Addgene, also known as clone 601 (Fig. 4.5) which contains the resistance gene for Ampicillin antibiotic. This resistence allows the selection of bacteria which have incorporated the plasmid with the Widom sequence by transformation.

A polymerase chain reaction (PCR) was performed with the proper primers designed in order to duplicate the Widom sequence. We included at the extreme of the forward and reverse oligos respectively cleavage sites for the Not I and Pst I enzymes. The sequences of the

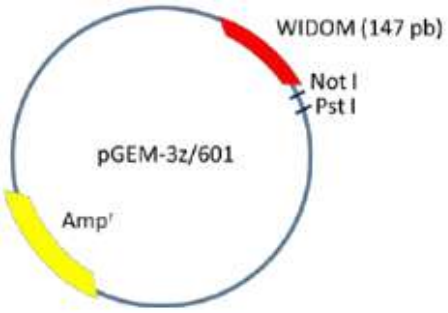


Fig. 4.5 Schematic view of pGEM-3z/601 plasmid which contains widom sequence, resistence gene for Ampicillin antibiotic and multiple restriction enzyme cleavage sites, emphasizing the unique cut sites of Not I and Pst I

direct oligo is 5'-AAGGAAAAAAGCGGCCGCACAGGATGTATATATCTGACACG-3' and the reverse 5'-AAAACTGCAGCTGGAGAATCCCGGTGCCG-3'. After the PCR reaction using the plasmid pGEM-3z/601 as a DNA template and obtaining the original WIDOM sequence with the restriction enzymes NotI ans PstI at the ends 5' and 3' respectively, the next step consisted in a double enzymatic digestion with NotI and PstI, both of the purified PCR fragment and plasmid pGEM-3z/601. Before that, the fragment obtained trough PCR reaction was purified with a purification column kit (DNA Purification Kit Gene Matrix Basic, EURx) after enough duplications, to separate the Widom sequence cloned from the remanent plasmid. After double enzymatic digestion and purification, a ligation process was carried out to bind the digested original plasmid (Vector) that already contains one Widom copy with the obtained fragment trough PCR product (insert) by T4 DNA ligase. Subsequently, the plasmids resulted in the ligation were introduced into competent *E. coli DH5α* cells.

CGCTGTTcaatacatgcACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGT
AATCCCCTTGGCGGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTG
CTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGgcggcc
gcACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCG
GTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTA
CGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGctgcagGCATG

Fig. 4.6 Sequence of the WIDOM-WIDOM fragment. It is shown in red the two WIDOM sequences and in lower case the NotI and PstI restriction sites. Nucleotides in capital letters in black at the ends are necessary for the specific amplification of the fragment and are Complementary to plasmid pGEM-3z/601.

We use competent cells as a method to select after ligation process those cells that have incorporated the plasmid which is correctly ligated, ie, as a method of transformant selection. The transformed cells were plated on LB agar plates supplemented with Ampicillin, which

is the antibiotic for which the vector confers resistance. After adequate growing time the plasmid is extracted from the cells by Miniprep (GeneJET Plasmid Miniprep Kit, Thermo Scientific). The final fragment including the two WIDOM sequences will be amplified with new specific oligos repeating the PCR and column purification step stated above. Sequence of this fragment is shown in Fig. 4.6.

The next step is the purification of the sequence of interest, which begins with an extraction of the nucleic acids by phenol/chloroform/isoamyl alcohol passes. Then, samples were dialyzed with GebaFlex Dialysis 6000-8000MWCO mega tubes and concentrated with Amicon Ultra-15 Centrifugal Filter 3 kDa in order to obtain the DNA in a smaller volume. Finally, using the Prep-Cell Kit (See Fig. 4.7) the sequence is purified by selecting the fragment of appropriate size, we quantify the eluted DNA and concentrate it again with the aid of the Amicon.

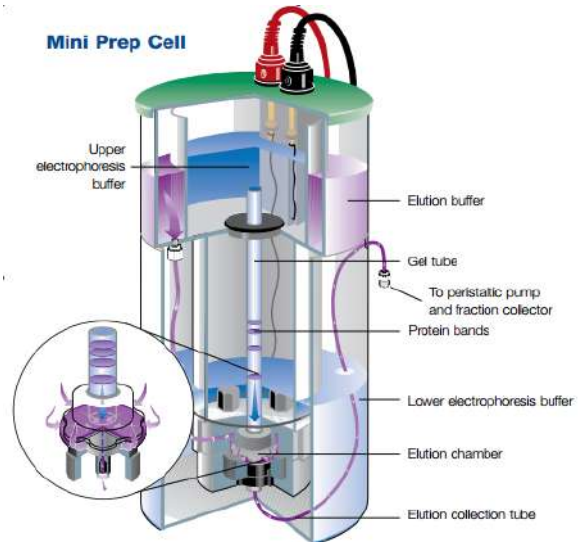


Fig. 4.7 Bio-Rad Prep Cell System. The system separates nucleic acids by electrophoresis of continuous elution, separating the DNA fragments by size into different liquid fractions.

The samples for SAXS and SANS experiments were prepared in a D_2O buffer containing 10mM Tris (pH 7.4), 0.1 mM EDTA and 50mM NaCl with a DNA concentration of 2.2 mg/mL. Two solutions, one with DNA and buffer and one with buffer only, were prepared for small angle scattering measurements. The samples were degassed in a partial vacuum of 0.5 bar for 3 hours before being loaded into quartz containers for measurement. The quartz cells for SANS were rectangular with a 2 mm thickness. Quartz cylindrical capillaries with an inner diameter of 1mm and a wall thickness of $10\mu m$ were used for SAXS.

Uses of Widom sequence in experiments

The combination of SANS and SAXS data proved to be particularly fruitful. The measured signal was small, particularly for the SANS data even despite the considerable amount of artificial DNA needed for the experiments, and using the two methods gave independent and complementary data.

SAXS gave a larger signal from much less sample, however x-rays can cause damage to DNA such as strand breaking. SANS does not investigate the atomistic structure of the particles, but is very sensitive to the global distribution of particle shape and size with linear dispersions of $1 - 100nm$. As a result, the combination of SANS and SAXS allows us to reveal some specific features of the "601" strong positioning sequence that complete the view deduced from the structural investigations of the nucleosome.

In the dilute scattering limit, the shape of a molecule in solution is related to a radius of gyration and a molecular form factor. As the temperature increases, dramatic changes in local stiffness are expected until the entire molecule eventually denatures in the melting transition. Knowing the sequence, we can apply our theory to calculate rigorously the persistence length as a function of position along the molecule and hence the radius of gyration and molecular form factor. These may then be compared to the SANS data, thus giving spatial information. SANS has been previously used on long-chain genomic DNA to determine persistence lengths [88], thus proving that the experiments are feasible.

In previous experiments carried out by our work team (TEST-2217, 9-13-467, 9-13-592, 9-13-625), we applied this methodology to the Widom-601 sequence and we recorded SANS curves and in-situ U-Vis spectra as a function of temperature throughout the transition. A specialized sample environment was constructed to enable the melting transition to be followed using in-situ UV absorption spectroscopy. We detected a clear signal related with the DNA that indeed changed with temperature and we were able to monitor the fraction of opened base pairs at any time using the U-Vis signal. The experiment was a great success from a technical point of view, however data analysis showed numerous issues with the choice of sample. The widom sequence was chosen because of its biological importance, however its length (145 base pairs, or 36nm) is short with respect to the persistence length (50nm). Changes in the persistence length thus resulted in small changes in the shape of the molecule, resulting in subtle changes in the scattering that were difficult to analyze quantitatively.

Despite the aforementioned problems with the sample, enough data was collected to defend qualitatively the hypothesis that the persistence length is sequence dependent. Fig. 4.8 shows UV absorption data. The data show a clear change with temperature. The absoprtion is proportional to the number of open base pairs and reflect the expected melting of the sample.

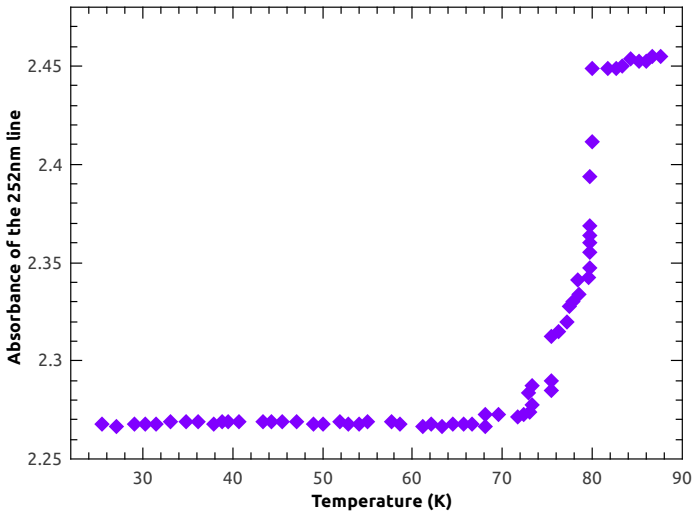


Fig. 4.8 In-situ UV absorption data measured during the data collection. The melting transition is apparent from the shape increase at 78 °C.

Seeing this promising results, we did believe that SANS was capable of performing such a measurement. Consequently, we wanted to repeat the experiment under optimal conditions. To that end, the first goal was to increase the length of the sequence under study to around the persistence length 50nm. The changes in the shape of the molecule would therefore be more readily observable. To that end, we used the sample whose preparation was explained above.

The SAXS measurements were performed using the VAXTER instrument at the Friedrich-Alexander Universität, Germany while the SANS measurements were performed using the D22 instrument at the Institut Laue-Langevin, France (9-13-665 experiment).A picture of the main elements of D22 instrument could be seen in Fig. 4.9.

With both techniques, SANS and SAXS, we calculate P(r) which is the pair-distribution function. It must be weighted by the appropriate scattering lengths for the radiation used to measure the cross-section, and it represents the ensemble average of the shapes of the particles in the solution. Mathematically, P(r) can be obtained from an inverse Fourier transform of the cross-section. We have used the ATSAS package [89] and particularly the GNOM program [90, 91] to calculate P(r) in this work. Figure 4.10-a) shows the measured SAXS and SANS intensities, I(q), at room temperature, and the corresponding P(r) are shown in Fig. 4.10-b).

The P(r) determined by the two techniques are qualitatively similar, but show significant differences, particularly for r$<$100Å. Both data sets show a peak at r$\approx$20Å. The peak in the x-ray data is due to the phosphates, which are on the outer side of the double helix and are 20Åapart. DNA contains many protons distributed on various sites of its structure and for

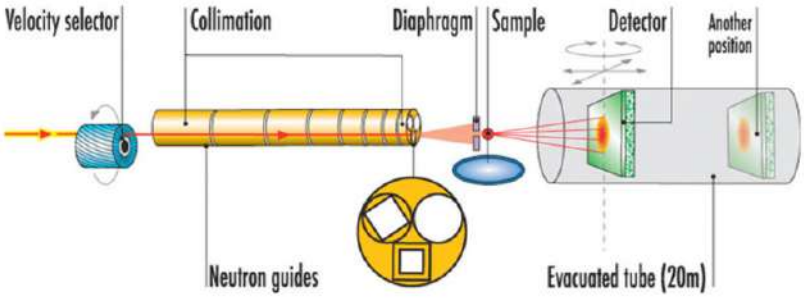


Fig. 4.9 Layout of the main elements (labeled) of the D22 instrument at the ILL (Grenoble).

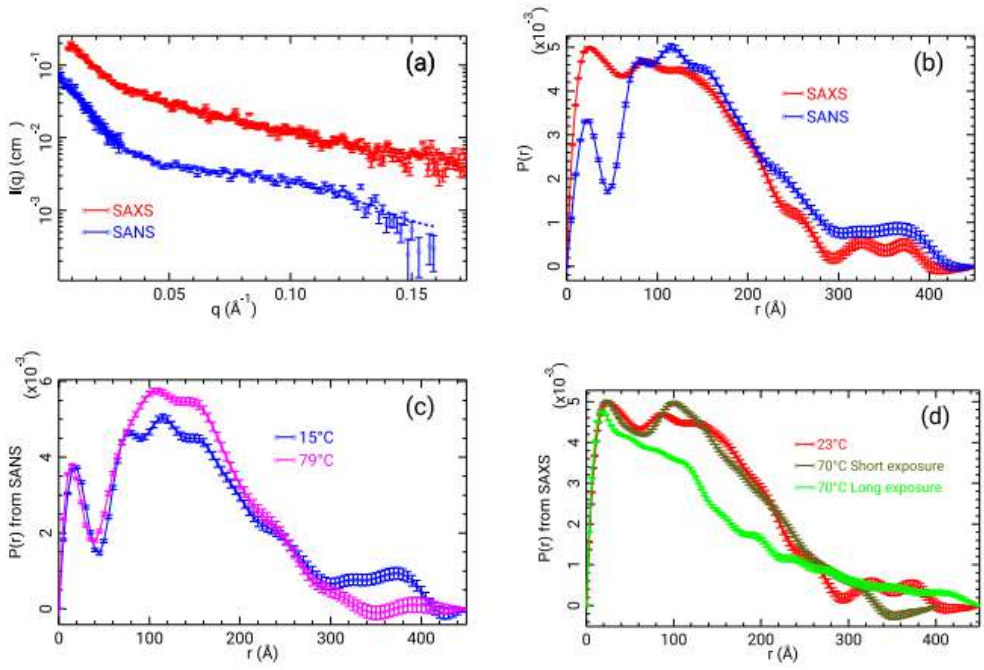


Fig. 4.10 (a) SANS and SAXS data measured at 25 °C and (b) their corresponding P(r). (c) the P(r) from SANS at 20 °C and 70 °C. (d) the P(r) from SAXS at 25 °C and at 70 °C as a function of exposure to x-ray radiation. The short exposure data result from a subsequent 1hr measurement after the sample temperature was raised directly to 70 °C. The long exposure data result from the final measurement of a sequence of 3hr run at 30 °C, 50 °C and 70 °C.

neutron scattering in short distances these is well approximated by the scattering by a bulk cylinder with a diameter of 20Å. Therefore, for both SANS and SAXS, the 20Åpeak appears to be the signature of the diameter of the double helix.

A simple polymer model for DNA was used to analyze the P(r) in Fig. 4.10-b)-d). The model, consisting of $N+1$ objects representing base pairs, each one separated by a fixed distance of a=3.34Åbut with bond angles that may vary at each base pair, is schematized in Fig. 4.11. It was an extension of the Kratky-Porod model [92], which itself is the discrete version of the worm-like chain model generally used for long DNA molecules. It was not concerned about the internal structure of the DNA, and was instead focused only on the conformations that the backbones could adopt. It is convenient to model using this method because the structure factor can be computed exactly, even when the bending energy varies along the polymer [93].

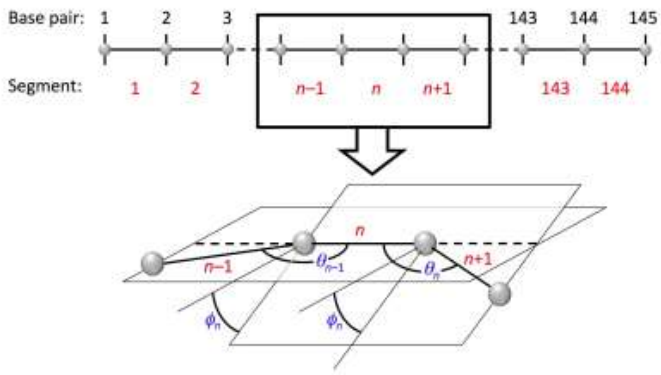


Fig. 4.11 Schematic picture of the polymer model used for the WIDOM-601 DNA measured in the experiments. The top part of the figure shows the definition for the numbering and the lower part shows the definition of the bond and torsional rotation angles.

Once a model conformation was determined, its P(r) was calculated by putting a unit scattering center at each base-pair position. The calculated P(r) were then scaled to have the same integrated area as the data. The P(r) were compared to the SAXS data at 25 °C which have better signal-to-noise than those for SANS.

The information in P(r) is not sufficient to systematically determine the geometry of the DNA molecules. Instead we scanned the conformational space by generating a large number of random conformations and we selected among them those which provided the best matching with the observed P(r). In total, $12 \cdot 10^7$ conformations were generated. The P(r) were calculated for those conformations accepted by the Monte-Carlo and were compared to the SAXS P(r) at 25 °C. The average P(r) for these conformations is plotted in Fig. 4.12.

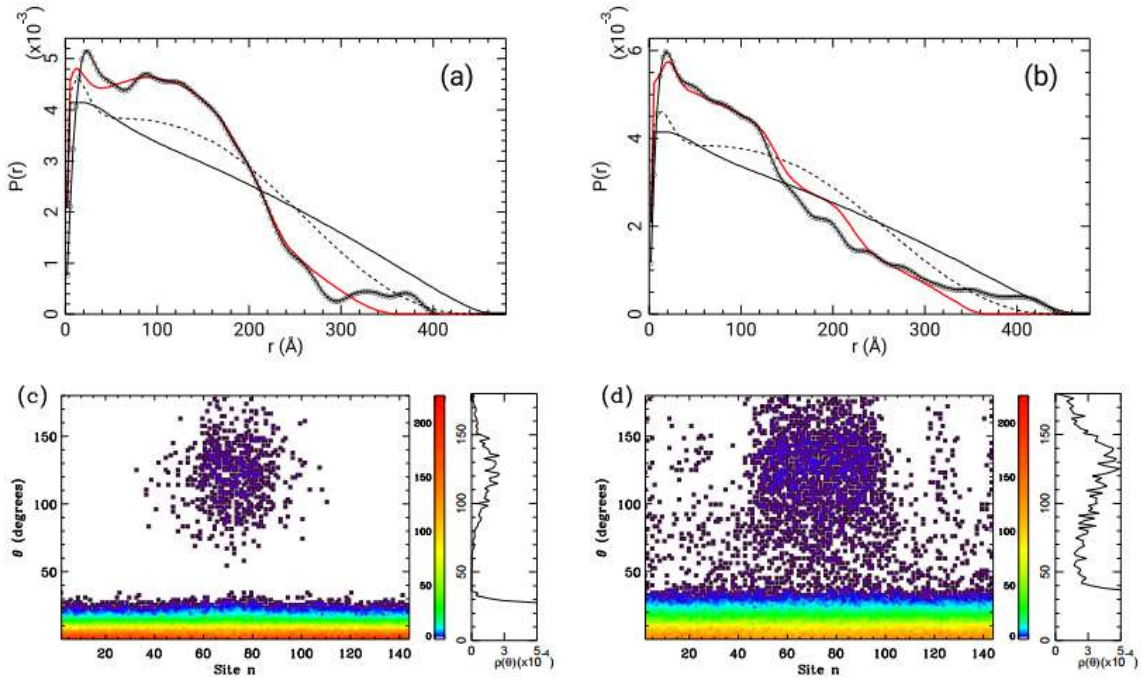


Fig. 4.12 (a) Comparison of the SAXS P(r) (black circles) with a model calculation for a homogeneous polymer, with the average P(r)(red) from the 10^3 conformations with the best match to the experimental data from the $12 \cdot 10^7$ conformations generated by Monte-Carlo. The full black line shows P(r) for a homogeneous Kratky-Porod model having a persistence length of 500Å. The dashed black line shows P(r) for the same model if one takes into account the torsional rigidity. (b) Shows the same as (a) for SAXS measurements after a long exposure to x-rays. (c) and (d) are the histogram of the bending angle against n for the 10^3 conformations that provide the best matching with the room temperature SAXS data.

The agreement is very good in the region 80Å$\leq$r$\leq$ 280Å. The calculation does not adequately capture the peak at r$\sim$20Å, however this is unsurprising as the model is insensitive to the diameter of the DNA which is the likely cause of the peak. The calculation neither adequately captures the plateau from 300Å$\leq$r$\leq$ 400Å. The 10^3 selected conformations were then examined to find common features characterizing the shape of the DNA molecules in solution and to design a model for these molecules. The conformational search showed that the molecules can exhibit a very sharp bend, that Crick and Klug [94] called as kink and they suggested is perfectly compatible with the structure and bonding of DNA. Within our polymer model the best agreement with experiments is obtained for a kink angle of 95^o, in perfect agreement with the kink predicted by Crick an Klug [94].

Therefore, the results strongly indicate the existence of kinked DNA molecules. They also suggest that, on its own, an enhanced local flexibility of DNA near the center of the "601" nucleosome positioning sequence is not sufficient to explain the small-angle scattering data, and that preferred non-zero bending angles must contribute.

Chapter 5

RAMAN Spectroscopy for the study of DNA melting

5.1 Experimental Techniques

5.1.1 Raman Spectroscopy

Raman spectroscopy is a spectroscopic technique used in condensed matter physics and also in chemistry for the study of vibrational, rotational and other low frequency modes in a system. It is based on phenomena of inelastic scattering (or Raman) of a photon of monochromatic light, usually through the beam of a laser in the spectrum of the visible or ultraviolet light, by its interaction with matter (Fig. 5.1).

The laser light interacts with the electron clouds of the molecules of the system, polarizing the cloud which surrounds the nucleus and causing the electrons to be excited, leading them to a virtual excited state whose energy is determined by the energy of the incident beam. These virtual states provide information about the vibrational modes of the system under analysis. That is to say, the vibrational and rotational frequencies of the molecules can be studied with Raman spectroscopy as well as infrared spectroscopy (IR). While these techniques are related to each other, the two types of spectra are not exact duplicates, but provide complementary information [95, 96].

Returning to Raman spectroscopy, the laser beam hits the sample and the electromagnetic radiation of the illuminated point is reflected by a lens and passed through a monochromator. When the virtual excited state decays, because it is unstable, the photons emitted (scattered light) have contributions of two types: the so-called elastic dispersion or Rayleigh, where the emitted photons have the same energy as the absorbed photons ($h \cdot \nu_0$), and is the one used, for example, in obtaining X-ray diffractograms; and the very weak inelastic dispersion or

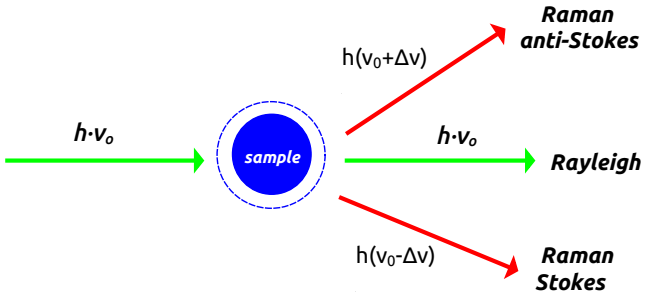


Fig. 5.1 Raman scattering effect.

Raman scattering ($\sim 10^{-5} - 10^{-8}$ of the incident beam) where the emitted photons undergo a change of energy after dispersion. This is what is known as the Raman effect in honor of its discoverer, the Indian physicist Sir C.V. Raman [97]. The Raman scattering may be accompanied by a reduction in the frequency (energy) of the scattered radiation, known as Stoke ($\nu_0 - \Delta\nu$), or may result in an increase in the frequency of the scattered radiation, anti-Stoke scatter ($\nu_0 + \Delta\nu$) [98]. In both types of Raman scattering, the energy of the emerging photons is not arbitrary, but must satisfy the relationship

$$\Delta\nu = \nu_0 \pm \nu_i \tag{5.1}$$

being ν_i the frequency of some of the vibrational or rotational modes of the material with which the photon interacts.

In Fig. 5.2 the elastic and inelastic scattering phenomena are represented, as well as the corresponding intensity ratios. In Stoke scattering, molecules promote from the fundamental vibrational state to a virtual state and finally decay to an excited state of vibration. As for the anti-Stoke dispersion, the molecules promote from an excited state of vibration to a virtual state and finally decay to the state of fundamental vibration. The Stokes dispersion has greater intensity than the anti-stokes scattering bands due to its density of excited states. This is so because of the Maxwell-Boltzman's distribution law. At room temperature, the population of molecules at a low vibrational level (ground state) is higher and the transition is more likely, therefore the Stokes bands are used to analyze the Raman spectra.

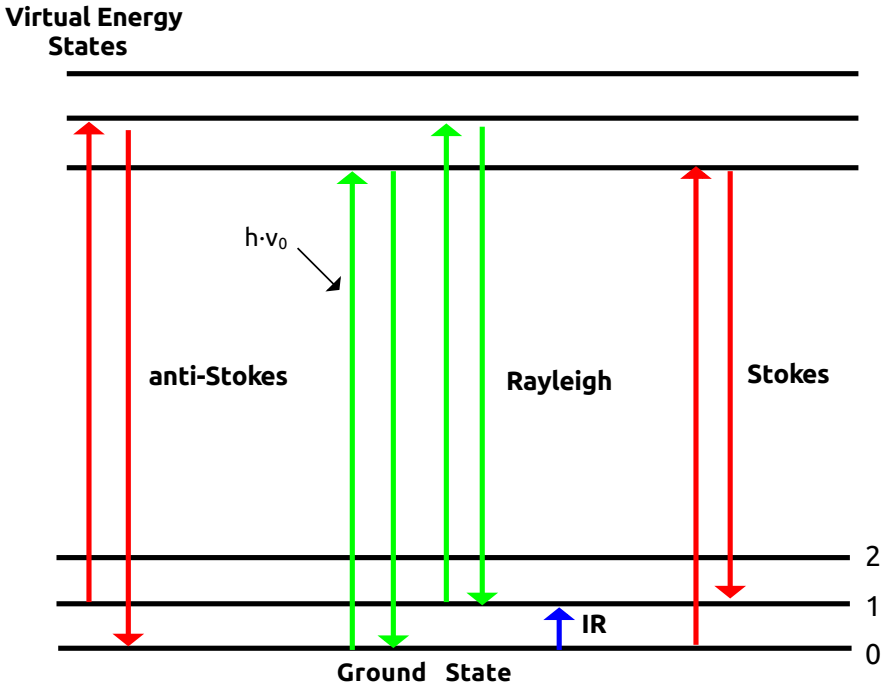


Fig. 5.2 Energy-level diagram showing the states involved in Raman spectra

As the temperature rises, the intensity of the anti-Stoke bands increases since so does the population of molecules at higher vibrational levels. The difference between the final and initial state is called the Raman shift and marks the position of the active Raman bands in the spectra [99].

Normally Rayleigh radiation is filtered, while the rest of the reflected light is scattered to a detector. But spontaneous Raman scattering is typically very weak, and as a result the main difficulty of Raman spectroscopy is separating the weak inelastically scattered light from the intense Rayleigh scattered laser light.

Only certain states of vibration are observable by Raman spectroscopy due to the rules imposed by quantum mechanics. A basic selection rule dictates that the most intense bands in Raman come from vibrations that produce a change in the dipole moment of the molecule (they measure the deformability of the bond due to the action of an electric field). Thus, symmetric vibrations, in which the facility to distort the bond depends on the distance that separates the nuclei, give rise to intense bands in Raman [100].

The Raman spectrum collects these phenomena by representing the scattered optical intensity as a function of the normalized wave number at which it is produced. The wave number is a quantity proportional to the frequency and inversely proportional to the wavelength, which is expressed in cm^{-1}

$$\overline{\nu} = \frac{1}{\lambda}, \qquad \overline{\nu} = \frac{\nu}{c} \tag{5.2}$$

where λ is the wavelength of the incident light, ν the frequency of light scattered by Raman effect and c is the speed of light. The Stoke and anti-stoke bands are placed symmetrically with respect to the Rayleigh scattering (See Fig. 5.3) (of great intensity and with the frequency of the incident photon), but as we have said previously, the Stokes bands are the ones used to analyze the spectra, placing the center of the Rayleigh band as the origin of the axis.

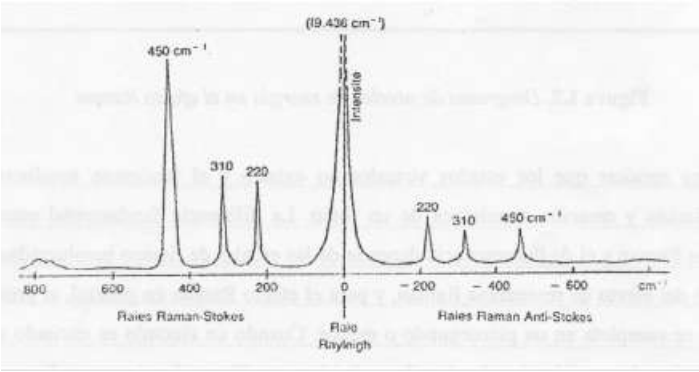


Fig. 5.3 Raman spectrum showing the Rayleigh band, at $0cm^{-1}$ in the center, on the left Raman-Stokes band, and to the right Raman Anti-Stokes.

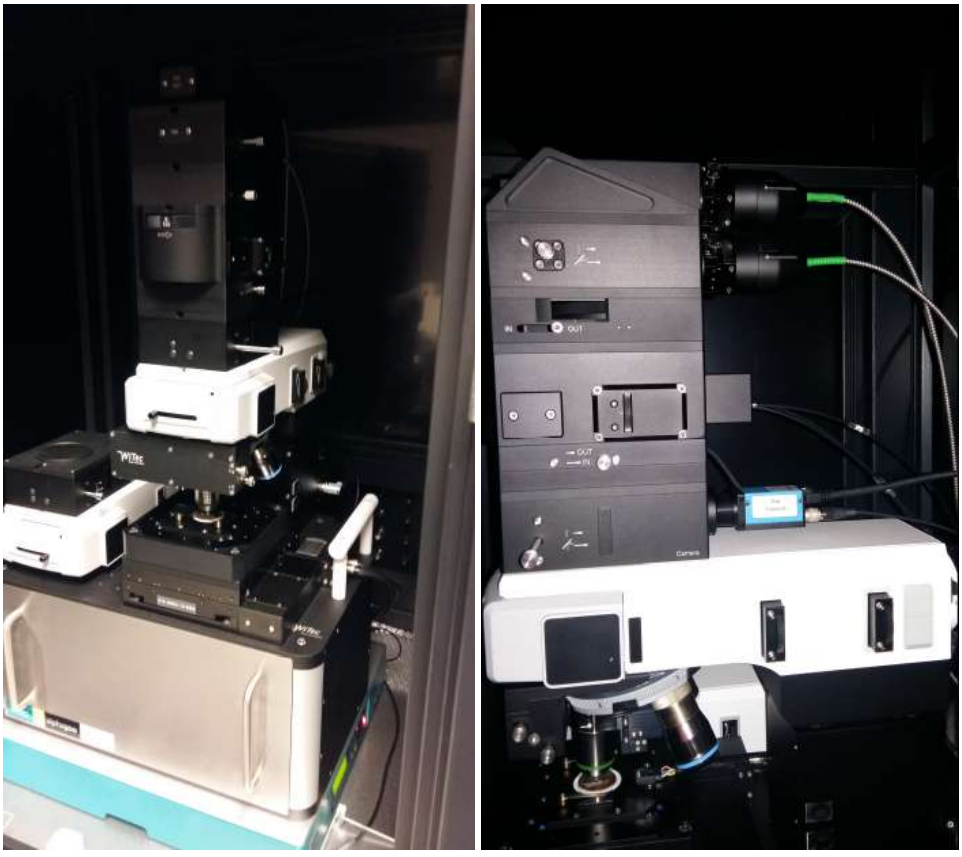
The source of excitement in Raman, which is usually a laser, is also important. Depending on the excitation energy there will be more or less dispersion according to the material we are analyzing, but the Raman line will be obtained exactly in the same wave number difference. That is, the wavelength of the emission source does not influence the displacement of the band (shift), but its intensity. The microscope with which we have carried out the measurements has 3 excitation sources available, lasers of 532nm, 633nm, 785nm. For our measurements we use the $532nm\ Nd:Y_3Al_5O_{12}$ laser.

Raman spectroscopy is a non-destructive, fast technique, which does not require an elaborate preparation of the samples and has a high spatial resolution, reaching it to be less than a micron. The Raman effect is used to infer different properties of a material, both chemical and structural. With this technique the chemical composition of organic and inorganic compounds can be analyze as well as materials in any state, solid, liquid and gaseous. It also provides information on electron-photon interactions, making it very sensitive to the crystallographic and electronic structure of the sample. It has been used

in research in very diverse fields, ranging from physics and engineering, to chemistry and biology.

In chemistry, apart from the composition, this technique is highly selective, since it allows obtaining vibrational information that is very specific of the bonds of the molecules. It provides a fingerprint of the molecule that allows it to be identified.

We will use Raman spectroscopy to study the bands present in DNA in fiber form, and we will analyze the behavior with the temperature of the intensity and the shift of these bands. That is, we will measure the melting transition of DNA fiber, paying attention to the resonance information of the Raman frequencies of the DNA base pairs. Raman measurements were carried out using a confocal micro-RAMAN *Alpha* 300 *R* instrument (Witec) (See Fig. 5.4)



(a) Frontal view (b) Side view

Fig. 5.4 Witec Confocal micro-Raman *Alpha* 300 *R* instrument.

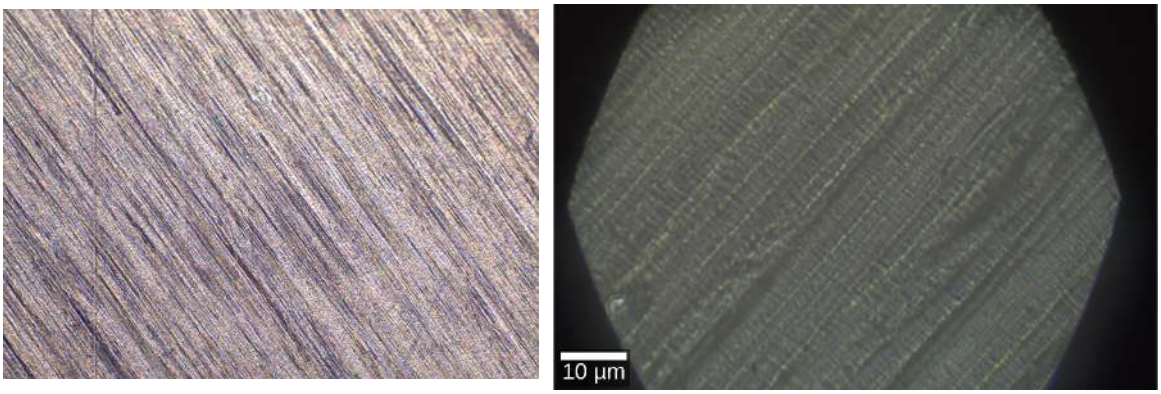
The Raman spectroscopy samples were introduced in a borosilicate glass capillary from Kimax® designed for melting point experiments and with both ends sealed. Each Raman spectrum was obtained with an integration time of 30*sec* and 10 accumulations. For the melting experiments, the temperature was rising up with a ramp of 0.5 - 1 °C per minute and

stabilizing each temperature during 10 minutes. Laser power used to capture the spectrums was $20mW$.

In addition, to correctly understand the results of the Raman, we will make a more detailed physical characterization of the fibers by means of an enhanced microscopy.

5.1.2 Optical Microscopy (OM)

We start the physical characterization of the DNA fiber by optical microscopy. Observations of this technique provides information about the fiber structure of the film. We recorded the image of the B-form DNA fiber sample with an *Olympus* optical microscope and a $100X$ objective lens. In Fig. 5.5 we can observe clearly the orientation of the fibers which forms the film and this allow us to make Raman measurements in different directions, parallel and perpendicular to the fiber axis orientation, in order to compare the homogeneity of the sample.



(a) 20X objective lens (b) 100X objective lens

Fig. 5.5 Microscope view of fiber obtained with Olympus optical microscope.

5.1.3 Atomic Force Microscopy (AFM)

We also characterize the same Na-DNA in B form fiber by Atomic Force Microscopy (AFM). The AFM is one kind of scanning probe microscopes (SPM). SPMs are a broad set of superficial analysis techniques at the microscopic level that emerged in 1981 in response to the limitations in resolution presented by optical microscopy. These techniques are designed to measure local properties, such as height, friction, magnetism, with a probe. To acquire an image, a sequential sweep is performed over a small area of the sample, recording the changes in the probe-sample interaction and measuring the local property simultaneously.

The interaction between probe and surface is only measurable when both are really close, making it essential to use a system that allows both to approach a distance within this range of action ($\sim 10\mu m$). Atomic force microscopy was one of the first techniques of this family to emerge in 1986 [101, 102].

The atomic force microscope is based in the detection of the strength produced between the atoms sited in the extreme of the probe and the atom or molecules that compound the surface of the sample. The kind of forces can be both long-range and short-range, as well as attractive or repulsive, including electrostatic forces, van der Waals forces or frictional forces. These forces exist in all materials, so the AFM is in principle applicable to any type of material. The nature and intensity of the forces exerted depend to a large extent on the distance (r) between tip of the probe and sample, and can be adequately described using a Lennard-Jones type potential

$$V(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{-12} - \left(\frac{\sigma}{r}\right)^{-6} \right] \tag{5.3}$$

here ε represents the height of the potential and σ the distance (finite) at which the potential is zero. This potential can be seen graphically in the Fig. 5.6.

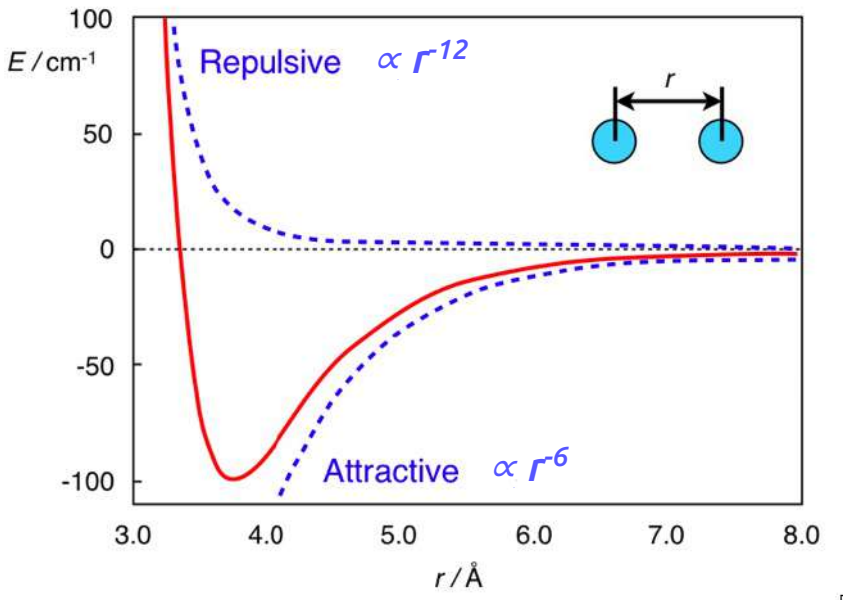


Fig. 5.6 Lennard-Jones potential. It allows to express the intermolecular forces between tip and sample at a function of the distance between them.

In order to quantify and control the magnitude of the interaction forces established between sample and tip, the latter is placed on the end of a flexible lever, denominated cantilever. The forces generated on the tip modify the behavior of the cantilever which means variations in deflection, oscillation amplitude, etc of the same. The changes induced over the cantilever are detected trough a laser that after incising over the cantilever, is reflected to a segmented photodetector. An outline of this process is depicted in Fig. 5.7.

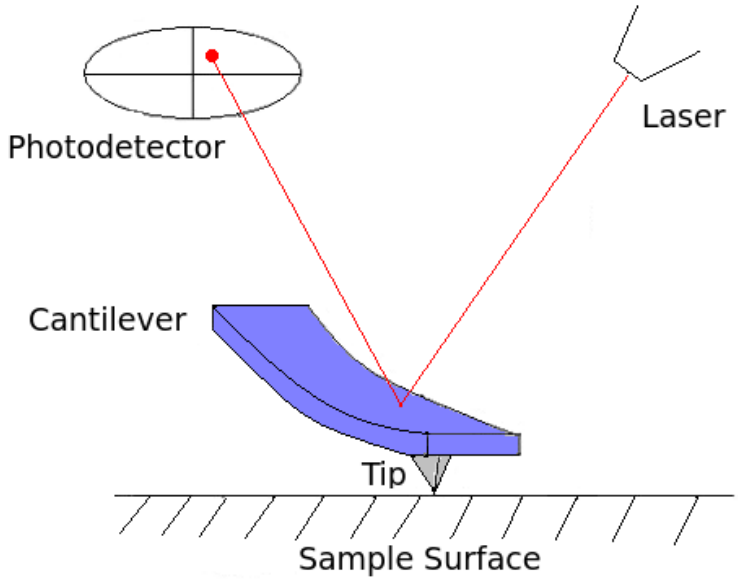


Fig. 5.7 Block diagram of atomic force microscope using beam deflection detection.

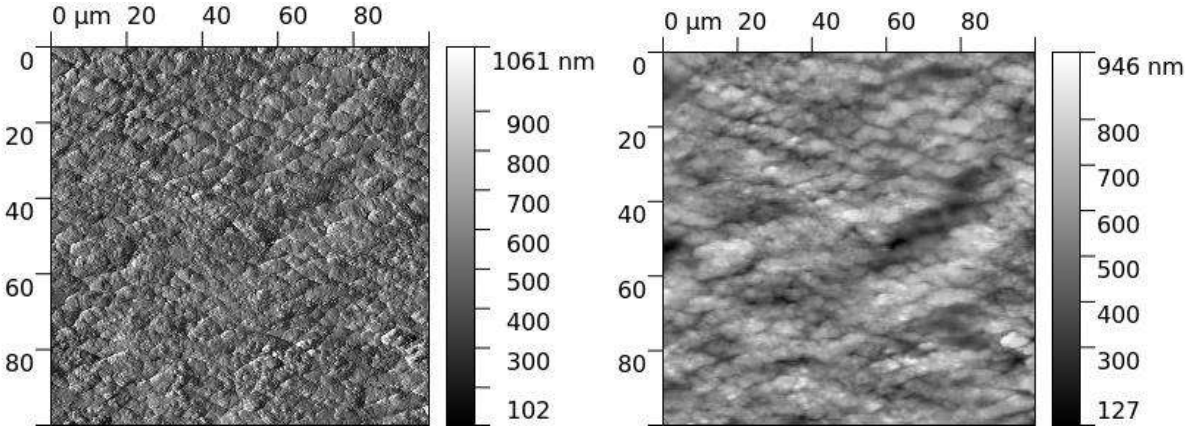
Taking into account the dynamics of the cantilever and what kind of interaction is established between tip and sample, there are differents operation modes of AFM [103, 104]:

- Static Mode: where the interaction force is detected following the changes in bending or torsion of the cantilever. It is known as a contact mode and can cause damage to samples.
- Dynamic Mode: there is no continuous contact between tip and sample since the cantilever is coupled to a piezoelectric oscillator forcing it to oscillate at a certain frequency. Within these modes we can cyte the modulation mode of amplitude or 'tapping' which is the one we use in our measurements.

The pictures of AFM showed in this work, were taken in tapping mode, using a silicon cantilever with a probe with a constant force of $1.2-6.4N/m(47-76KHz)$ in the CSI Nanoobserver equipment showed in Fig. 5.8.

Fig. 5.8 CSI Nanoobserver AFM equipment.

Amplitude and topography images of AFM (Fig. 5.9) show a succession of regions which give rise to fiber orientation being each region a zone where DNA is compacted and could be formed by chains intertwined with some orientation.



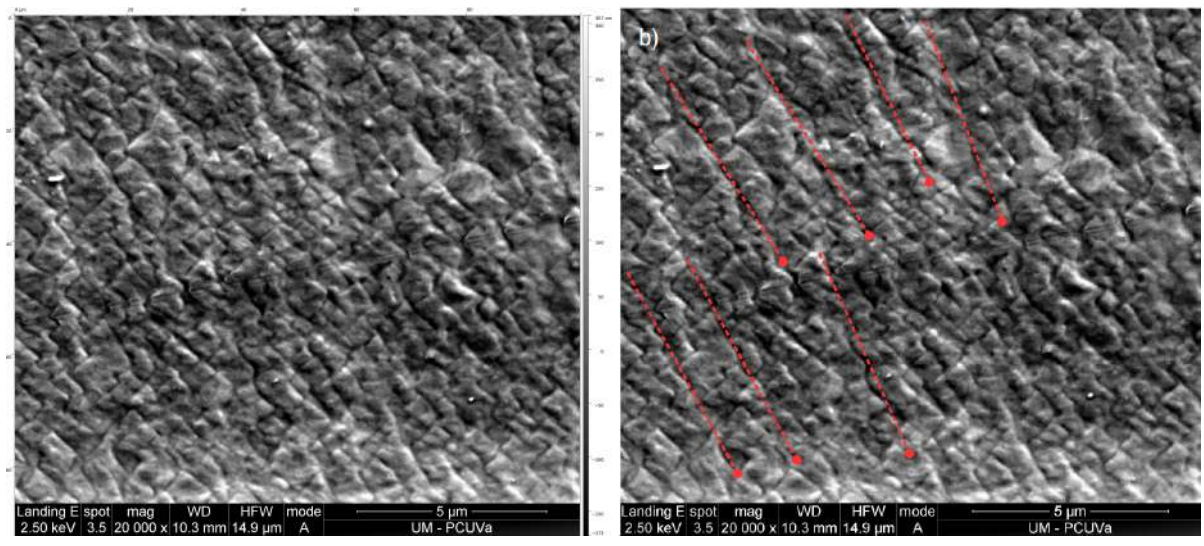
(a) Amplitude (b) Topography

Fig. 5.9 Atomic Force Microscopy image obtained with a CSI-Nanoobserver equipment of a Na-DNA fiber. Left panel shows the amplitude picture while right panel is the topography image.

5.1.4 Scanning Electron Microscopy (SEM)

Another form of characterization of the fibers was by scanning electron microscopy (SEM). Using an electronic microscope we can get images of the morphology of the sample regardless of its thickness, with a resolution of $\sim 1-50nm$. It is usually used as a substitute for the optical microscope when its resolution is not sufficient, although in our case we have complemented both techniques. From an SEM, information on the topography of the surface can be extracted from the backscattered electrons or by secondary electrons that give a contrast image between regions of different composition. The backscattered electrons are a fraction of the electrons of the incident beam that leave the material without having lost part of their energy in inelastic scattering processes. They also give information on the composition, because the heavier atoms generate more backscattered electrons. In turn, the secondary electrons are produced as a consequence of inelastic processes of ionization of the incident beam. This type of electrons is quickly reabsorbed by the atoms of the material due to its low energy (<50eV) so that the more superficial atoms contribute more to form the image, which is more faithful to the surface topography [105, 106].

For DNA fibers, the images were taken at the microscopy unit of the scientific park of the University of Valladolid. SEM pictures were recorded with ESEM FEIQuanta 200 FEG microscope using the beam deceleration mode and a typical image is shown in Fig. 5.10.



(a) (b)

Fig. 5.10 Scanning Electron Microscopy (SEM) picture of DNA fiber using the beam deceleration mode (BMD) with a Microscope $(ESEM)\ FEI-Quanta\ 200FEG$. Red dashed lines in right panel indicate fiber orientation.

Seeing Fig. 5.10, it can be observed more clearly the filaments which form the fiber, a completely different texture and even some kind of vertical orientation (red dashed lines) that suggest the arrangement of DNA chains.

5.2 DNA Fiber analysis by Raman specroscopy

In previous chapters, we already talked about the importance of oriented samples for structural studies. For example, DNA structure was elucidated by Watson and Crick [13] with a X-Ray image obtained with a DNA fiber. We also explained the main role of DNA thermal denaturation and DNA bubbles for many biological processes. And how the knowledge of DNA melting transition is still a challenge. We also have seen that our working team has been the first one in reporting structural information through the melting transition of B-DNA fibers by means of neutron diffraction, measuring the spatial correlation along the molecule, and achieving in this regard a success where other techniques failed [20, 40, 107, 108]. In addition, studies of DNA melting have been mostly investigated in solution by techniques such as UV absorbance, circular dichroism and calorimetry. In this chapter, we are going to study DNA denaturation transition in fiber form by Raman spectroscopy, paying attention to the resonance information of the Raman frequencies of the DNA base pairs. We will also study the melting transition of DNA in solution in order to find differences in the melting transition and in the behavior with the same processes in fiber state, where DNA has some structural restrictions because stacking and solution. Finally, we will do a comparative study of DNA fibers but in presence of different solvents (and different kinds of fiber, Na-DNA and Li-DNA) analyzing the influence of the solvent in the behavior of the fiber.

DNA melting is observable in fiber and gives favorable agreement between theory and experiment. We analyze the temperature dependence of the intensity and the shift of the bands present in the DNA film Raman spectrum and correlate this behavior with the melting transition and with the structural changes that accompany denaturation of DNA films.

Earlier works dealed with investigate DNA thermal denaturation in solution by Raman Spectroscopy. Therefore, relying on bibliography we are able to assign the main bands present in DNA Raman spectrums measured at room temperature. As we mentioned above, samples consisting in DNA fibers obtained with 'wet-spun' technique [73–75] were introduced in a borosilicate glass capillary from Kimax® and with both ends sealed (See Fig. 5.11. The room temperature Raman spectrum was obtained with an integration time of 30sec and 10 accumulations, with a laser power of 20mW.

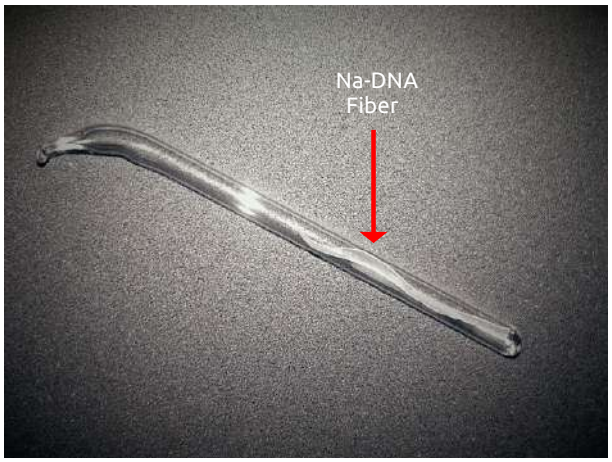


Fig. 5.11 Raman sample consisting in DNA fiber introduced in a glass capillary with both ends sealed. Hereinafter, we will refer this kind of sample as dry fiber, since it is not exposed to any solvent or special atmosphere.

Fig. 5.12 shows the DNA spectrum recorded at 20 °C and we can see that most of the classical bands have been identified in these spectra. Based on bibliography we assign the main bands in the spectrum as it can be shown in Fig. 5.12. For easier localization, we listed the bands that suffer significantly changes during the melting in both intensity and shift in Table 5.1.

The bands are the following: $727cm^{-1}$ corresponding to ν dA and which intensity proportional to unstacking of dA [109–112]; $782cm^{-1}$ assigned to backbone $O-P-O$ of B-form and C2' and C3' endo of dC [109, 112, 113]; $803cm^{-1}$ which is a marker of $O-P-O$ and can vary with DNA form and suffer a shift in the premelting [111, 113]; $1012cm^{-1}$ ν_{CO} *st* desoxirribose ring [109, 114] ; $1100cm^{-1}$ assigned to PO_2 symmetric stretch A-form [109, 115, 116]; $1183cm^{-1}$ which corresponds to marker dT y dC, thymine unpairing (paired/unpaired state) [109, 111–113] ; $1248cm^{-1}$ is also a marker of DNA form and C3' endo anti of dC. It suffers a hypochromism and high change in intensity with melting. It is also assigned to scissoring y wagging dA C1'-N, N1=C6 and C6-N6 exo dA [109, 110, 112–115] ; $1334cm^{-1}$ assigned to C2' endo anti and N7-C5 dG, C3' endo anti and C8=N7 dA and experiment a shift with removing hydrogen bonds [109, 112, 113, 115]; $1372cm^{-1}$ very remarkable C4=O of dT which suffers a strong change and it has a frequency shift with removing hydrogen bonds [109, 111, 112]; $1482cm^{-1}$ assigned to N9-C8 dA and dG ring bands, decreasing the band frequency with temperature as N7 can form hydrogen

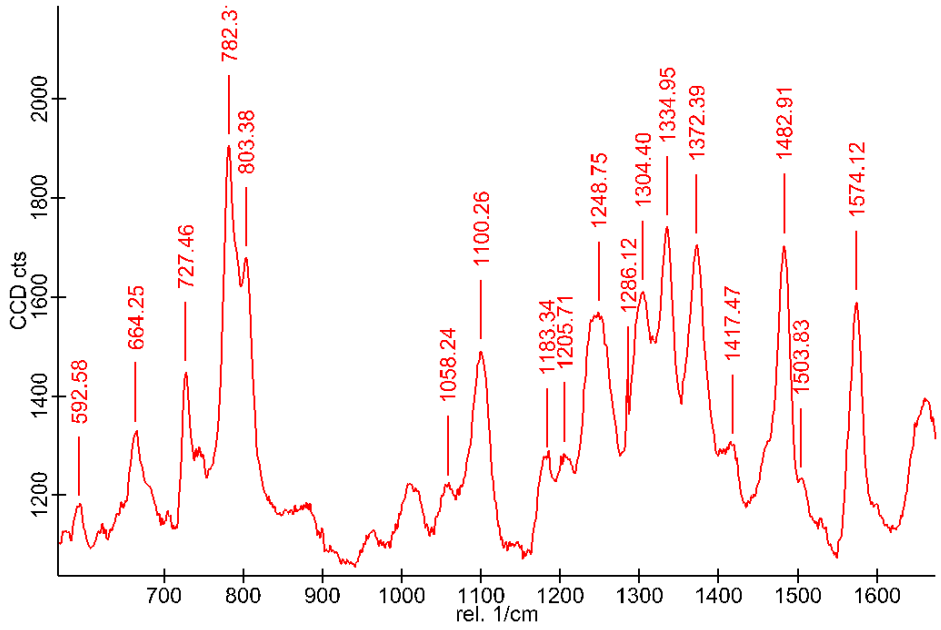


Fig. 5.12 Spectrum of DNA recorded at 20 °C showing main and characterized bands.

Table 5.1 Wavelength (cm^{-1}) of the bands present at Raman spectrum of Fig. 5.12 for DNA dry fiber.

Band Position (cm^{-1})	**Assignation**
664	marker dG $C3'$ endo syn
727	ν dA
782	O-P-O B-form; $C3'$ and $C2'$ endo dC
803	O-P-O marker
1012	ν_{CO} st desoxirribose ring
1100	Po_2 symmetric stretch A-form
1183	markers dT, dC (paired state)
1248	marker DNA form;$C3'$ endo anti dC
1334	$C2'$, N7-C5 dG; $C3'$, C8=N7 dA
1372	C4=O dT
1482	N9-C8 dA and dG
1574	N6-H2 ring mode dA and dG
1660	C=O, C5=C6 N-H dG, dC and dT

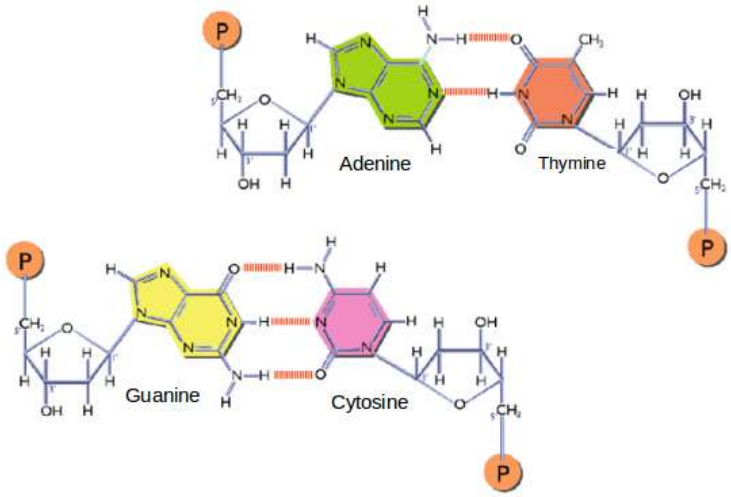


Fig. 5.13 Schematic representation of DNA nucleotides. The vibration of some of this bonds generate the bands presents in DNA Raman spectrum.

bonds [109–111, 115] ; 1574cm^{-1} that corresponds to ring mode of dA and dG (N6-H2 ring *st*) and it is typical in AT rich DNA [109, 111–113, 115] and 1660cm^{-1} assigned to ν_s C=O, C5=C6 and N-H of dG, dC and dT and whose intensity increase a lot with with temperature [109, 111, 114, 115].

These bands correspond mostly to the vibration modes of the bonds of the nucleotides that form the DNA. To make it easier to visualize and see which bond corresponds to each band, Fig. 5.13 is a schematic representation of these nucleotides.

We repeated several measures in different places of the same dry fiber in order to verify the homogeneity of the sample. As it can be observed in Fig. 5.14, where we perform the record of the Raman spectra along a line parallel and perpendicular to the fiber axis orientation, the resulting spectrum are really similar, seeing that the differences between them (left down panel) are not relevant.

Hence, due to homogeneity of the sample, we can corroborate that DNA films are an excellent model to study some DNA properties as melting transition.

The study of the melting transition in DNA fibers is the aim of this chapter, and the way followed is the analysis of the bands present in the Raman spectra at different temperatures. In the spectrums, we analyze two different parameters: Intensity, which give an idea of thermal broadening of each group of vibrational mode and unstacking of the bases (Raman hypochromism), and the shift of the band frequency which suggests elimination or formation of hydrogen bondings and is related with changes in deoxyribose-phosphate backbone [110, 111]. The shifts in Raman band frequencies change due to localized structural perturbations.

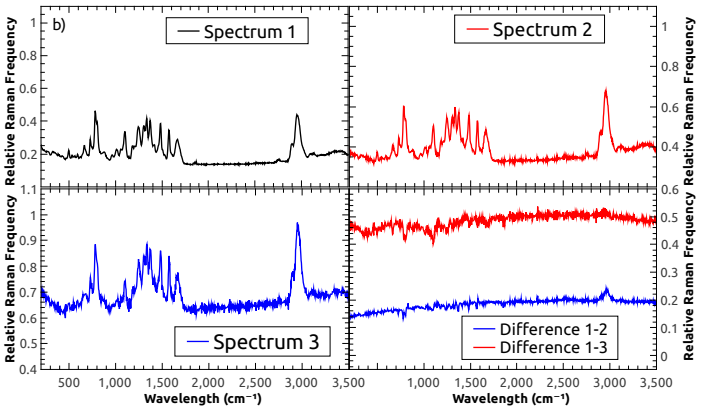


Fig. 5.14 Raman spectrums recored at different parts of the same fiber to check the homogeneity of the sample. First 200 wavelength has been removed form the spectrum in order to avoid Rayleigh scattering.

Increases in the intensities of many Raman bands are also observed during the melting, reflecting both unstacking and unpairing of bases attendant with strand separation [112]. An example of the changes that occurs in the Raman spectrum when the temperature is rising up, is depicted at Fig. 5.15

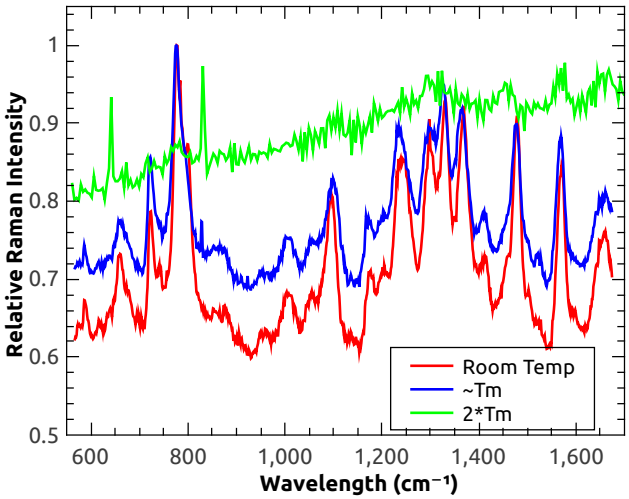


Fig. 5.15 Raman spectrum of DNA fiber at room temperature (red), around T_m (blue) and really above melting temperature (green) comparing changes in the intensity and frequency shift of bands.

In this picture (Fig. 5.15) Raman spectra of DNA dry fiber at room temperature (20 °C), around melting temperature (T_m) and really above it, two times T_m, where the fiber is almost fully denaturated are shown. The aforementioned changes in the intensity and frequency shift

of the bands are cleary observed. In order to compute the changes occurred in both, intensity and frequency of the bands while the temperature is rising up, we fitted all the recorded spectra with a Lorentzian equation(5.4):

$$f(x) = y_0 + \frac{2 \cdot A_i}{\pi} * \frac{w_i}{(4 \cdot (x - x_i)^2 + w_i^2)} \tag{5.4}$$

where A_i is the integrated intensity of each peak of the spectrum, w_i is the full width of the band at half height and x_i is the center position of each peak. With this parameters we are able to follow the changes in the intensity and in the frequency of the bands presents in DNA spectra during melting transition.

5.3 Melting transition of DNA fiber studied with Raman spectroscopy

In the previous section we have shown the main bands present in DNA fiber Raman spectrum (See Table 5.1). We want to study denaturation transition so we focus in the behavior with temperature of these bands with more biological sense or more structural relevance. Following the changes occurring in the bands with temperature we will correlate these changes with structural changes going on in the fiber during the denaturation.

The canonical way to study thermal DNA melting is the use of denaturation curves, so we decided to generate melting profiles of each band, identified in the bibliography like structurally relevant for DNA. We based our curves on the evolution of the intensity and the frequency shift, as previous works made in solution [111, 112, 117, 118]. We started by analyzing changes in the intensity of specific bands as a function of temperature. The melting curves generated with these changes of the specific bands of DNA dry fiber as function of temperature are shown in Fig. 5.16.

For comparing spectral intensities we follow the next criterion: normalization relative to the maximum intensity of the Raman line near $1100 cm^{-1}$ in each spectrum. The intensity of this line is changed only slightly with temperature and its half-width is invariant too. Moreover, the $1100 cm^{-1}$ line is well-separated from near Raman lines and its integrated intensity, as we saw before, can therefore be easily measured.

We can see that as the temperature is increased from room temperature to one hundred degrees, the intensity of all Raman modes, except $780 cm^{-1}$, increases substantially. This increase, is a quantitative measurement of some effects in DNA chain and indicates melting transition. It is exhibited for hypochromic bands of DNA with the thermal denaturation [111].

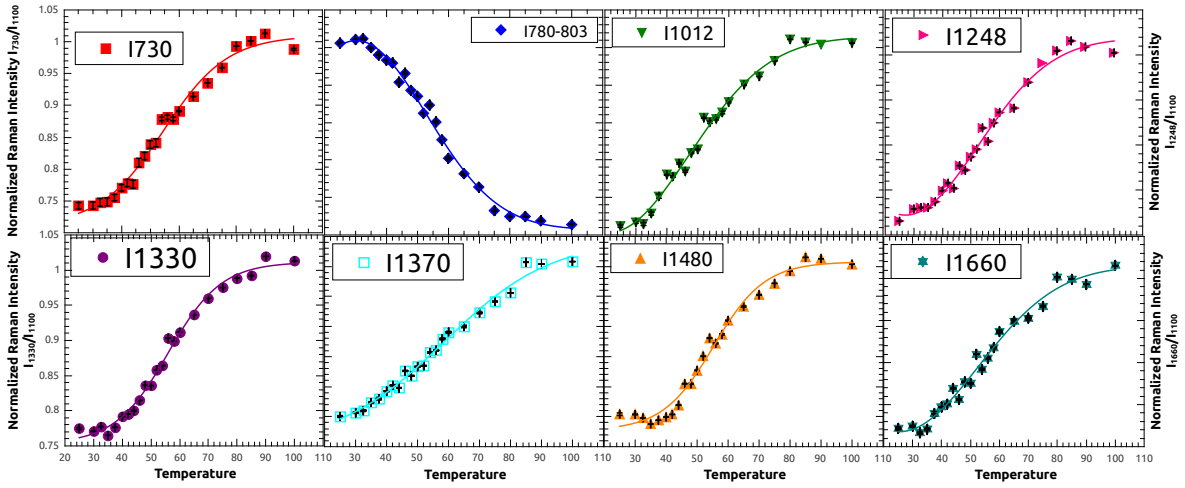


Fig. 5.16 Temperature-dependent relative Raman intensities of the $730, 780, 1012, 1248 cm^{-1}$ (upper panels) and $1330, 1370, 1480, 1660 cm^{-1}$ (down panels) modes of DNA fiber. All intensities that are shown represent an average which correspond to three independent replicates and have been normalized to their maximum value. Error bars are shown for each data.

The increase in Raman intensity is proportional to unstacking and unpairing of bases during melting transition.

Now we will address the meaning of the intensity increase in each band. First curve on the left panel in Fig. 5.16 corresponds to the band at $730 cm^{-1}$ which is assigned to unstacking of dA. It can be observed that the melting profile of this band occurs at lower temperature and faster and the transition is narrower than in the case of the other bands. This makes sense since adenine residues (which only has two hydrogen bonds between base pairs) denature before. Following the order, next curve is for the phosphodiester marker at $780 cm^{-1}$ where we can see a decrease in the intensity as DNA undergoes thermal denaturation. When DNA is denaturated the backbone present higher conformational flexibility so it that is the reason why the intensity decrease. Moreover, this band merges with the one at $803 cm^{-1}$ and we computed the combined intensity of the two. $803 cm^{-1}$ line is also a marker of $O-P-O$ and is well established to corresponds to A-DNA form. As we mentioned in the section of fiber preparation, the films are obtained in a humidity atmosphere of 75% and in this conditions DNA is mostly in A-form (with some B contaminations) [119] so Raman peaks corresponding to A-DNA is not surprising. But there is a band at $790 cm^{-1}$ reported in B-form, so the combination of these two bands is good choice and reflect the conformational mixture of the samples. Band at $1012 cm^{-1}$ is assigned to ν_{CO} *st* desoxirribose ring [109, 114] so is related with the backbone of the molecule. Therefore, as mentioned above, as backbone has higher conformational flexibility when denaturation occurred, it

is reflexcted in an icrease of the intensity of this band. The last panel on the left in the first row of Fig. 5.16 is for the band $1248cm^{-1}$ which is also assigned to dA residues so it has the same behavior in the intensity increment mentioned before for $730cm^{-1}$ band. The first panel below corresponds to $1330cm^{-1}$ mode which mainly arise from stretching vibrations of imidazole ring [120] thus is attributed to changes in dG and dA residues. Next one is the curve of the $1370cm^{-1}$ band, related to the out-of-phase stretches of thymine ring bonds coupled to C4=O [115, 120]. The changes in the intensity of this curve could be coupled once again with thymine denaturation. It is particularly interesting the Raman band near $1480cm^{-1}$, which is mainly assigned to the guanine ring. Residues of guanine have one site, N7, which is really sensitive to interaction with electrophiles. Consequently, the changes here can be attributed to a different environment of this group. Finally, the broad band centered near $1668cm^{1}$ is the sum of two individual peaks ($1662cm^{1}$ and $1693cm^{1}$) and contains contributions from coupled C=O stretching and NH deformation modes of dT, dG, and dC, with dT as the major contributor so the denaturation behavior it is mostly attributed to this nucleotide.

One conclusion that can be drawn in light of the results presented in Fig. 5.16 is that they demonstrate melting curves well defined and with enough statistics therefore are pretty good taking into account that represent the eight frequencies more significant to explain DNA structure. With the changes in these bands, denaturation transition can be followed easily.

Now it is time to pay attention to the information of the Raman frequency shift. We complete our analysis generating similar melting profiles for the bands of key importance in DNA structure, with the changes on the band shifts due to temperature increament.

We have already seen that the shifts of the bands in Raman spectrum is a parameter that suggests the elimination or formation of hydrogen bondings and that is related with changes in deoxyribose-phosphate backbone [110, 111]. The frequency shift also reflects differences in hydration associated with crystalline packing of the double helices [121, 122]. Thus, the bands that exhibit changes in frequency are those related to the possibility of forming hydrogen bondings once DNA denatures and its residues are more exposed. Fig. 5.17 shows denaturations curves obtained analyzing frequency shifts occurred in main bands of DNA dry fiber as a function of temperature. Moreover, we do an estimation of the calculation of the melting temperature and the width of the transition by fitting the generated denaturation curves, so we can compare this obtained data with previous works.

Shifts in Raman frequencies reflected in spectral changes are tipically related with highly localized structural perturbations. Raman hypochromism, which we have shown that is correlated with increase intensity during melting and can occur for bands assigned to pyrimidine and purine ring vibrations involving the concerted stretching of conjugated single

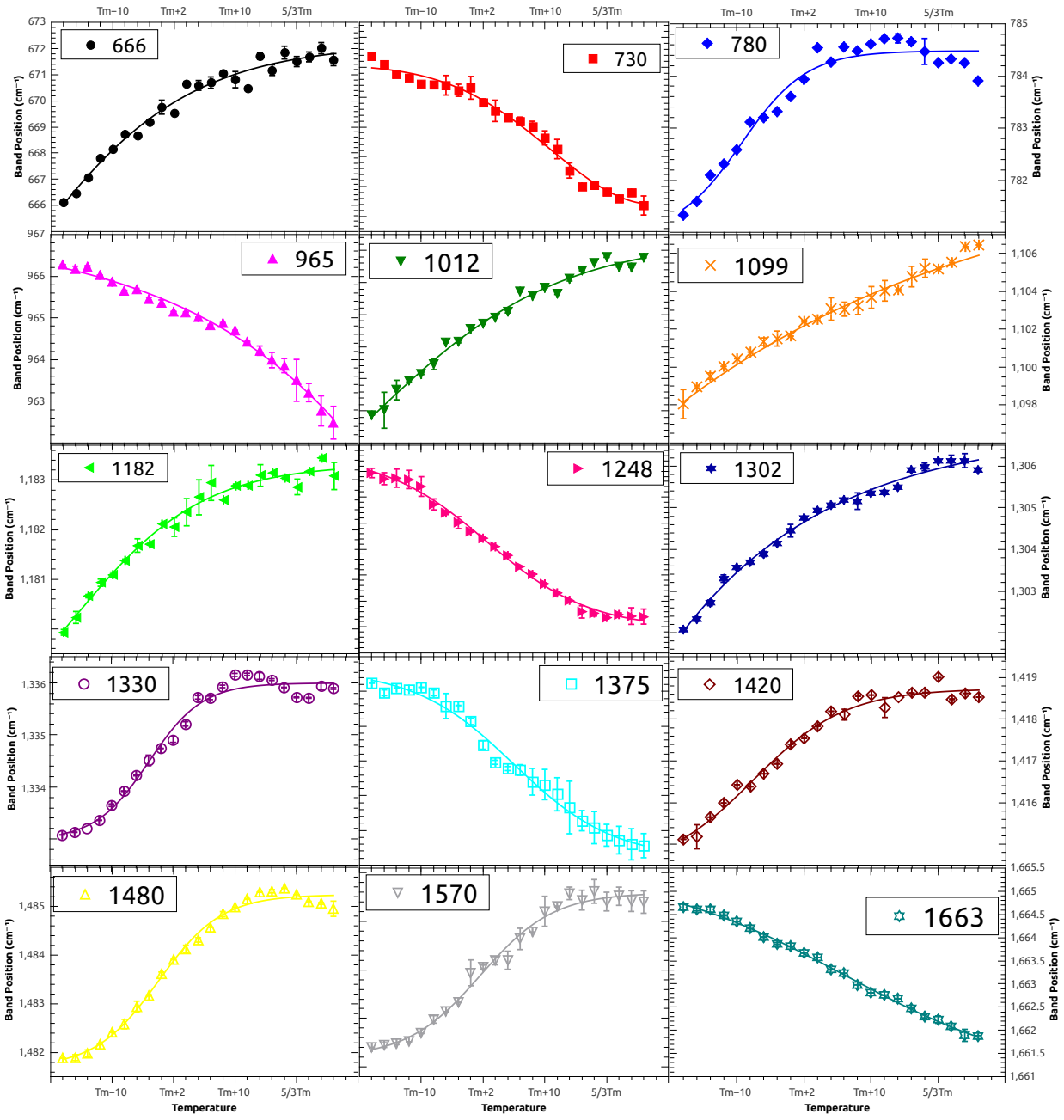


Fig. 5.17 Temperature-dependent Raman frequency shift of DNA dry fiber bands. Frequencies result from an average of three independent replicates. All error bars are shown.

and double bonds in the heterocycles, may also be accompanied by a detectable shift in the frequency of the band center [111]. This happens for example with the dA marker at $727cm^{-1}$. This band appears at $727cm^{-1}$ in A-DNA and at $729cm^{-1}$ in B-DNA. In DNA B-form, this band shifts from $735cm^{-1}$ to lower frequencies while gaining appreciable intensity upon melting of the native structure. With this shift one corroborate once again the conformational mixture present in our samples. Bands at $1240cm^{-1}$ and $1480cm^{-1}$ present however first criterion being the suffered shift due to a structural perturbation. In Fig. 5.17 other bands that exhibit frequency shift can be observed. Phosphodiester (OPO) stretching at $800cm^{-1}$, which we have seen that is diminished in intensity in direct proportion to the elimination of A DNA structure during denaturation [123], is barely discernible at high temperatures or converge with the band at 780 that suffer a shift to $790cm^{-1}$. The line near $965cm^{-1}$, which we can see because is more intense in A structures, is due to bond-stretching (CCand/or CO) in the deoxyribose ring [109], so is related with the backbone of the molecule and therefore is consistent the shift suffered for this peak with denaturation. Other example is the PO_2^- band at $1100cm^{-1}$ that suffers a significant wavenumber shift. Elimination of hydrongen bonding are also really tipycal in the shifts $1249 \rightarrow 1243cm^{-1}$ (dA), $1375 \rightarrow 1371cm^{-1}$ (dT), and $1573 \rightarrow 1576cm^{-1}$ (dA). Band at $1333cm^{-1}$ assigned to the stretching of C8=N7 and N7-C5 of adenine [120], is known to show a frequency shift of up to $4cm^{-1}$ if all of the hydrogen bonds disrupt at N1, N3, and N7. Stronger hydrogen bonding with the solvent can be attributed to $1420cm^{-1}$ band which shifts to upper frequency with thermal denaturation (guanine N7 site is exposed in denatured DNA) or purine band at $1573cm^{-1}$. And also, the broad band at $1668cm^{-1}$ which we already said is a sum of two different bands and comprises dT C4=O and C5=C6 stretching vibration, is sensitive to H-bonding too. Carbonyl stretching vibrations typically experience an increase in frequency with the elimination of hydrogen bonding.

Here we can conclude again that our results are in good agreement with the DNA melting theory. Melting profiles show the denaturation transition with properly shape and only differing in the width of the transition where some are sharper (indicating a cooperative strands separation) and other ones are smoother.

By fitting the obtained denaturation curves, we calculate the width of the melting transition. Melting temperature is also possible to calculate with this fittings but in this case we understand that it does not make much sense since T_m can be calculated precisely when the sequence is known. In our case the fibers are formed by an average of sequences so the melting temperature will be around 50º (very similar to our results). In addition, the experimental setup and the environment is very different from that reported by other papers with which we could compare the fiber denaturation temperature. In the case of our experimental setup,

we had temperature fluctuations. The lost of temperature during its transmission along the peltier, the sample and its support made the temperature being not exact, even though we had several ways to measure it.

To calculate the width of the transition is of importance since it gives information about the phase transition and its cooperativity. Furthermore, this data allow us to compare our Raman results with other well established techniques, as neutron scattering or calorimetry, that evaluate structural information of the fiber during the melting. Regarding the width of the transition previous publications suggest it occurs in a narrow temperature range, being narrow transitions more cooperative [11, 124]. One should to keep in mind that our sample is mostly A-DNA conformation so maybe the comparations could be not easy.

In order to calculate the width of the DNA fiber melting transition with our Raman results, we use different bands assigned to base pairs, backbone and sugar ring. The selected bands as well as its calculated width are listed on Table 5.2. The average width of the transition is 5.03 ± 1.44 suggests a smooth and continuous transition as reported previously [20, 107].

Table 5.2 Width of the melting transition calculated by fitting melting profiles of characteristic selected bands.

Type of band	**Position** (cm^{-1})	**Width**
Base pair	1248	4.12
	1370	4.15
	1420	6.3
Sugar ring	1335	4.26
	1480	5.03
	1570	5.27
Backbone	780	4.85
	965	3.23
	1012	8.13

From this study of the melting transition on DNA fiber with Raman spectroscopy some conclusion could be made. There are some specific bands in Raman spectrum of DNA that are helpful to identified DNA conformations and allow to discerns between A- and B- DNA form. Also, there are specific bands related with structural information and allow to follow and analyze the denaturation. For example, frequency shifts occurred in specific bands like $1240, 1370$ or $1660 cm^{-1}$ are related with the elimination or the formation of new hydrogen bonds between base pairs and solvent once DNA is opened. Also, structural changes suffered by the backbone, reflected by the increment of the intensity of bands such 730 or $800 cm^{-1}$, can be correlated with denaturation transition. Finally, the width of the transition calculated

with Raman data is in good agreement with previous studies that report structural information.

5.4 DNA fiber v.s DNA solution melting transition.

Once studied the melting transition in DNA fiber we also did an analysis of the denaturation of DNA in solution with Raman spectroscopy in order to compare both transitions and correlate them with structural changes produced in the double helix in both states, fiber and solution. We have seen that DNA Raman spectrum exhibits some characteristic bands assigned to specific vibrational modes of bases, sugar ring or backbone residues which are highly sensitive to thermal disordering [111]. Through the pages of the present thesis, we have made a lot of progress understanding how it works DNA fusion at the structural level. To continue this line, we now compare the changes occurred during the melting transition in ordered samples (DNA fibers) with the ones produced in DNA solution.

DNA fibers were made using the "spinning" technique [74, 75] from DNA sodium salt. The final DNA film was obtained after a re-humidification processes in a 75% relative humidity. This final process set the water content and stabilize the structure of the film. As we said in previous sections, samples were equilibrated to 75% RH which gives mostly A-form with some B form contamination [119, 125].

DNA solution was made with the same DNA sodium salt than in DNA fibers from salmon testes from Sigma-Aldrich company. The solution present a concentration of $30 mg/mL$ to ensure a good Raman signal and both samples, fiber and solution, were introduced in glass capillaries sealed. Again, Raman spectra were recorded with an integration time of $30 sec$, 10 accumulations and laser power of 20mW. In the melting, temperature was rising up with a ramp of 0.5 - 1 °C per minute and stabilizing each temperature during 10 minutes.

We listed the main bands present in DNA Raman spectrum on Table 5.1. Once we have main bands identified, we measure a Raman spectrum of both samples, fiber and solution, at room temperature. This result can be observed in Fig. 5.18.

Fig. 5.18 shows a clear difference between fiber and solution reflected in the frequency shift of some peaks in the spectrum. We already know the meaning of the frequency shift; it is a parameter that suggests elimination or formation of hydrogen bondings and it is related to changes in deoxyribose-phosphate backbone [110, 111]. In addition, frequency shift reflects differences in hydration associated with crystalline packing of the double helices [121, 122]. Thus, comparing the shift occurred in both DNA dry fiber and DNA solution, and between them, we will be able to study how it affects the transition in both states, and to check structural differences between fiber and solution.

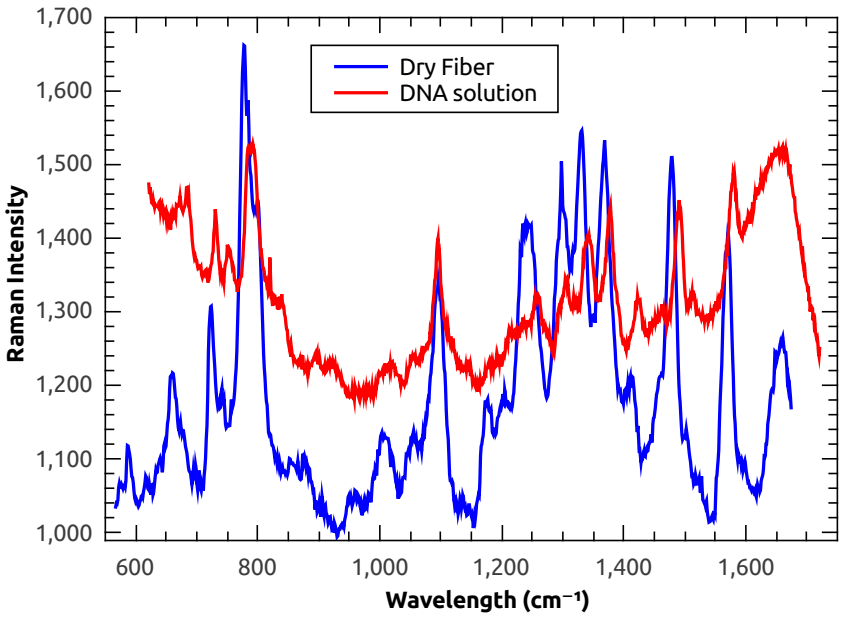


Fig. 5.18 Raman spectra of DNA solution (red) and DNA fiber (blue) measured at room temperature.

To localize easily the frequency shift differences between fiber and solution we collect the wavelength of the peaks of Raman's spectra of Fig. 5.18 of each sample in Table 5.3

Table 5.3 Wavelength (cm^{-1}) of the peaks of Raman spectra of Fig. 5.18 for DNA dry fiber and DNA solution samples.

Dry Fiber	**DNA solution**	$\Delta\sigma^2$
729.45	725.55	3.897
784.91	781.07	3.832
803.57	829.32	25.756
1008.24	1009.35	1.113
1249.4	1253.35	3.95
1334.34	1335.3	0.96
1373.05	1371.45	1.6
1488.46	1485.69	2.77
1575.97	1574.37	1.6
1662.91	1645.32	17.599

A new parameter is also shown on Table 5.3. We define three variables which are going to allow us to identify which frequencies in each state are changing more ($\Delta\sigma^1_{shift}$) , the structural differences between fiber and solution ($\Delta\sigma^2_{shift}$) and how the transition occurs in each sample and the differences of the denaturation between them. Eqs. 5.5-5.7 define this variables

$$\Delta\sigma^1_{shift}Fiber = |Band\ Fiber\ RT - \ Band\ Fiber\ T_i| \tag{5.5}$$

$$\Delta\sigma^1_{shift}Solution = |Band\ Sol.\ RT - \ Band\ Sol.\ T_i| \tag{5.6}$$

$$\Delta\sigma^2_{shift} = |Band\ T_i\ Sol. - \ Band\ T_i\ Fiber| \tag{5.7}$$

Paying attention to the results of Table 5.3 that summarizes results of spectra shown in Fig. 5.18, it can be observed that some frequencies of the main bands change and suffer a shift between solution and fiber at room temperature. But there are two bands that experiment a bigger change, bands $803cm^{-1}$ and $1662cm^{-1}$ respectively.

Band at $803cm^{-1}$ corresponds to phosphodiester stretching marker $O-P-O$ which suffers a different structural stress if it is in fiber or in solution. In DNA fiber the phosphodiester groups are more constrained while in solution, DNA has an increment in conformational flexibility of the backbone. This could explain the difference present in the $803cm^{-1}$ band between solution and fiber. Furthermore, it is known that phosphodiester marker facilitates the use of Raman spectroscopy to distinguish among DNA conformations A ($803cm^{-1}$), B ($830cm^{-1}$) and Z ($745cm^{-1}$) from one another [109, 111, 113, 126]. DNA in solution is always in B form (in Table 5.3 we can see the band at $829cm^{-1}$) but as we explained above fiber samples were stabilizes in a 75% relative humidity which gives a majority of A-form (DNA film is more dehydrated so there is A-form mainly present in the fiber) as indicate by the band at $803cm^{-1}$. The other significant shift between solution and fiber at room temperature it is present in the broad band centered $1662cm^{-1}$. We already know that this band is a sum of peaks attributed to stretching of $C_6 = O$ (dG), $C_2 = O$ (dC), $C_4 = O$ (dT), $C_5 = C_6$ (dC, dT) and N-H deformation with dT the major contributor[111, 113, 118]. It is also known that this band is highly sensitive solvent conditions and the disruption of Watson-Crick hydrogen bonding. Thus, the pronounced shift to lower frequencies of this band, that we see in solution ($1645cm^{-1}$), could reflect different contributions of thymine carbonyl acceptor when it is bonded with water molecules or adenine residues. Contrari-

wise, DNA fibers present structural restrictions due to the close-packed arrangement of the molecules so the carbonyl acceptors shows weaker hydrogen bonding in double helix and this band appears at higher frequencies [115, 126]. Some bibliography also reported that the frequency of the carbonyl group in B-DNA and Z-DNA structures differ significantly by virtue of a shift of the guanine $C_6 = O$ stretching mode to $1662cm^{-1}$ in the Z structure[126] suggesting that the fiber could contain also a mix of Z-form. This is not weird at all since Z-form can appear in genomic DNA rich in GC residues and the starting material for DNA films is DNA sodium salt from salmon testes with G-C content approximately of 40%.

Once we have seen main changes in the frequency shift at room temperature, we pass to study this changes but due to temperature incrementation in order to study the melting transition in both states and structural differences between them. We want to see how the transsition is affected in each state and correlate it with structural changes. For do that, we decided to analyze the evolution with the temperature of the parameters defined by Eqs. 5.5-5.7. Moreover, we repeat the generation of melting profiles (the tradicional way to study melting transition is by denaturation curves) based on the parameters that these equations suggest as relevant changes at a funtion of temperature.

This is what is shown by Figure 5.19. It depicts $\Delta\sigma^1_{shift}$ for solution and fiber and $\Delta\sigma^2_{shift}$, for the main bands presents in the DNA Raman spectrum at a function of temperature.

Fig. 5.19 shows the evolution between Room Temperature and a temperature really above transition, two times T_m going through intermediate states and the melting temperature. Focusing first in the evolution until melting temperature, it can be observed that three new frequencies are more affected at $0.8T_m$ besides the ones that we already have seen that changed at room temperature. There is an increment in $\Delta\sigma^2_{shift}$ for bands $729cm^{-1}$, $785cm^{-1}$ and $1249cm^{-1}$. The band at $729cm^{-1}$ is assigned to ν dA and is a marker of unstacking of this nucleotide [110–112] so makes sense that rising up the temperature, this band be affected. Adenine residues denature earlier because only has two hydrogen bond between base pairs. Fiber state presents more structural restriction than solution, making adenine residues less accessible, so the differences between DNA solution and DNA fiber are remarkable. In addition, some previous works inform that DNA A-form present this band at $730cm^{-1}$ while B-DNA has it around $725cm^{-1}$ [113] suggesting once again that DNA fiber is formed mostly by A-DNA. Next significant change is at $785cm^{-1}$ band which corresponds to the ring breathing of dC and is a diagnostic for base stacking in cytosine residues [113, 127]. Although cytosine has three hydrogen bond between base pairs, this band present the same behavior that the previous one. Cytosine residues are also more restricted in fiber state so the difference between fiber and solution with the temperature increase is evident. The last change produced at this temperature is at band $1249cm^{-1}$ that is a mixture of the adenine

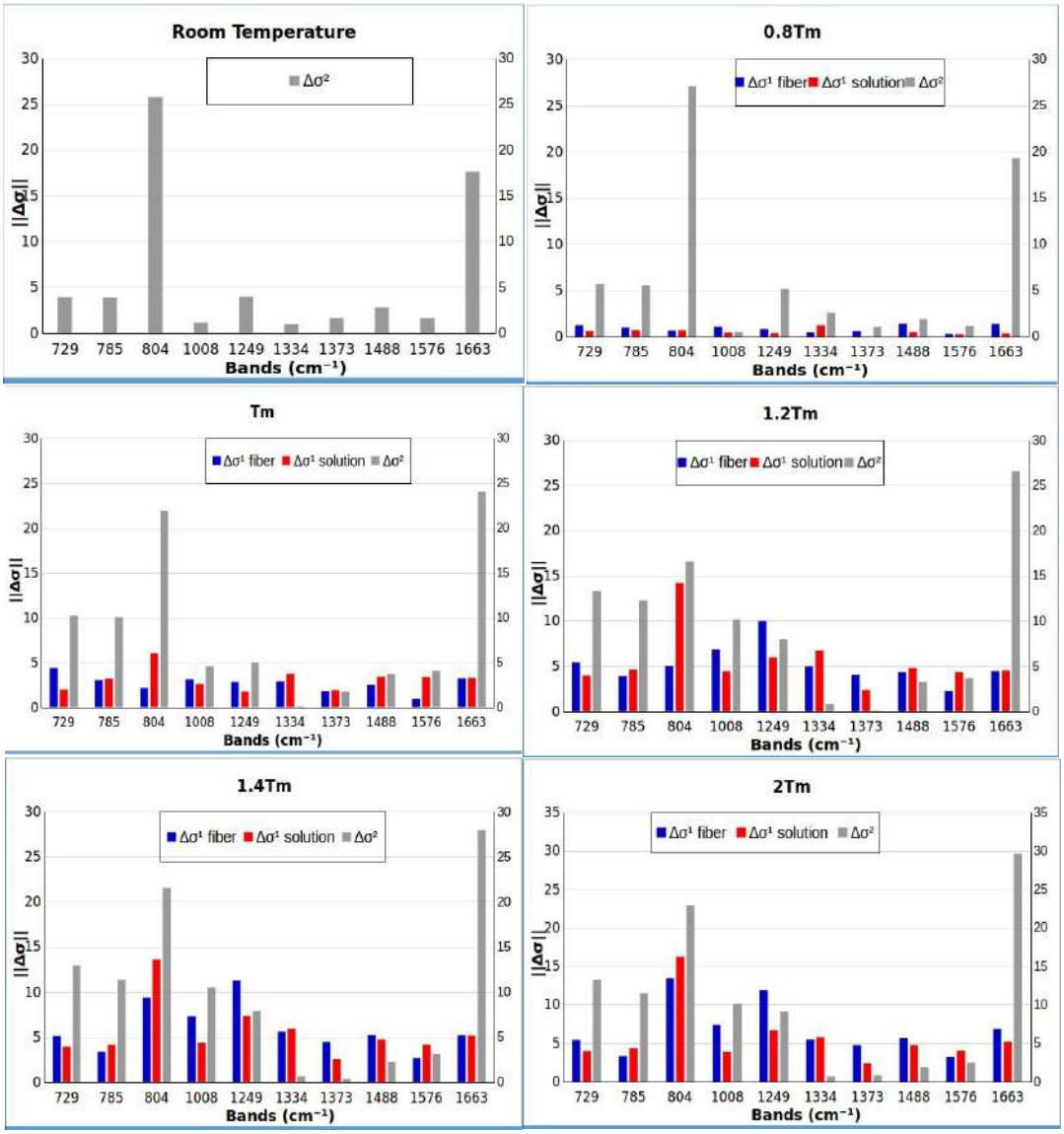


Fig. 5.19 Evolution of the parameters defined by Eqs. 5.4-5.6 with increasing temperature. These parameters represent the shift suffered by the main bands present in DNA Raman spectrum.

exocyclic C6-N6 bond and the N1=C6 ring bond and ring vibrations of thymine, being a marker of dT unstacking[110, 111, 115]. This band is also a marker of B-form [112, 114] therefore it is more affected in DNA solution making $\Delta\sigma^2_{shift}$ bigger since there is a difference between solution and fiber. When we are close to the melting transition, at T_m, it can be observed that more or less the same bands are remarkably affected but within them $729cm^{-1}$ has a higher change in fiber ($\Delta\sigma^1_{shift}Fiber$) while in solution the shift more altered is in $803cm^{-1}$ band ($\Delta\sigma^1_{shift}Sol.$). We already have commented that the band around $729cm^{-1}$ appears at different frequency depending on DNA form, and here it is clear that DNA fiber mostly formed by A-form has the band at $733cm^{-1}$ at T_m while DNA solution totally in B-form has this band at $723cm^{-1}$ making evident the differences between both states. In the case of DNA solution, is also clear that band at $803cm^{-1}$ is more affected that in fiber state because this marker also appears at different frequency according to DNA form but besides DNA in solution has bigger flexibility of the backbone letting phosphates fluctuate widely.

Once the melting temperature is exceeded, we start to see other bands that present significant changes between states and also inside each of them, besides the ones that was already affected whom keep increasing the differences. The more remarkable change between states reflected by parameters of Eqs. 5.5-5.7 is at $1008cm^{-1}$ band. This band corresponds to the sugar moiety, specifically stretch of $C-O$ desoxirribose ring. This backbone vibration do not show any premelting, it is seem to remain intact up to the melting point. The $1008cm^{-1}$ band is found to get boarder and weaker upon melting, specially for fiber where the center of the band shifts to higher frequencies contrary to DNA in solution where the band shifts to lower frequencies. Therefore, this band is an indication of differences between fiber and solution but it also suggests an important change in DNA fiber itself. There is other change that attract our attention because has the reverse trend. It is the band at $1249cm^{-1}$ upon melting temperature, where it can be shown that now, this band is more affected in fiber. One possible explication is that once the fiber is denatured adenine and thymine residues which denature first, are more expose and the difference inside the fiber state between room temperature and upon T_m is higher than in solution. Finally, it should be pointed out that other three bands present changes and these are more intense in DNA solution. This bands are bands at $1334cm^{-1}$, $1488cm^{-1}$ and $1576cm^{-1}$. The $1334cm^{-1}$ mode mainly arise from stretching vibrations of imidazole ring [120] thus is attributed to changes in dG and dA residues and experiment a shift with removing hydrogen bonds. Band near $1488cm^{-1}$ is also of particular interest being assigned mainly to the guanine ring and it is sensitive to interactions of electrophiles with N7 acceptor which could form stronger hydrogen bonding with solvent. Similar behavior is ovserbed for the purine band at $1576cm^{-1}$ that corresponds to ring mode of dA and dG (N_6-H_2st) [112, 115]. As we have seen, this three bands are well

related with elimination or formation of new hydrogen bonding so it is logic that are more affected in solution state. Opened DNA is in contact with the solvent whom can form this bonds and DNA strands have more flexibility since base pairs are now separated therefore they are free to fluctuate and to form new hydrogen bonds.

Until now, we have seen the bands that more change in each state, fiber and solution, correlating structural changes produced on them and how is affected the transition in each one. As we did before while we studied melting of DNA films, we generate melting profiles of the same bands for DNA solution and DNA fiber, with the bands that we just have seen are more significant. Furthermore, we fit obtained denaturation curves in order to calculate the width of the melting transition and compare it in both states and with previous results. In Fig. 5.20 we plot together generated melting curves based on the changes in the frequency shift of specific bands of DNA fiber and DNA solution, at function of temperature.

Firstly, it can be observed once again that the obtained melting curves exhibit a good agreement with canonical denaturation curves, showing properly shape and enough statistics. To generate this profiles, we have selected frequencies that are more significant upon to explain DNA structure based on the most striking changes in the parameters of the Eqs. 5.5-5.7. This bands also allows us to better follow denaturation transition.

In light of results observed in Figure 5.20 we also can conclude that exist differences in the melting transition between solution and fiber where fiber present in all cases a smoother transition while solution exhibits sharper transition indicating a cooperative strand separation. Even for some bands (1334 and $1576cm^{-1}$), both states present different trends but the same behavior. For example, for $1334cm^{-1}$ vibration mode, fiber state exhibit an increase of the frequency shift while DNA solution presents exactly the contrary having a decrease of the shift. The same it could be observed for the band at $1576cm^{-1}$, but in both cases transition width is really similar despite that the behavior of the bands with the temperature are different. The bands that are more similar regarding transition width and melting temperature in both states, fiber and solution, are the ones at 1249 and $1660cm^{-1}$, being the two of them related with denaturation of thymine residues. Regarding the other two frequencies present in Fig. 5.20, 729 and $803-830cm^{-1}$, we can see that DNA solution exhibit a narrower transition produced in a short temperature range while DNA fiber presents a clearly smoother transition, and the range of which is produced the frecuency shift is smaller.

Table 5.4 lists the width of the transition calculated for different bands in both states. We have obtained this wdth by fitting the generated denaturation curves and use it to evaluate structural melting transition of DNA fiber and melting signatures in DNA solution. We use representative bands assigned to base pairs, backbone and sugar ring to calculate the width of the DNA denaturation transition.

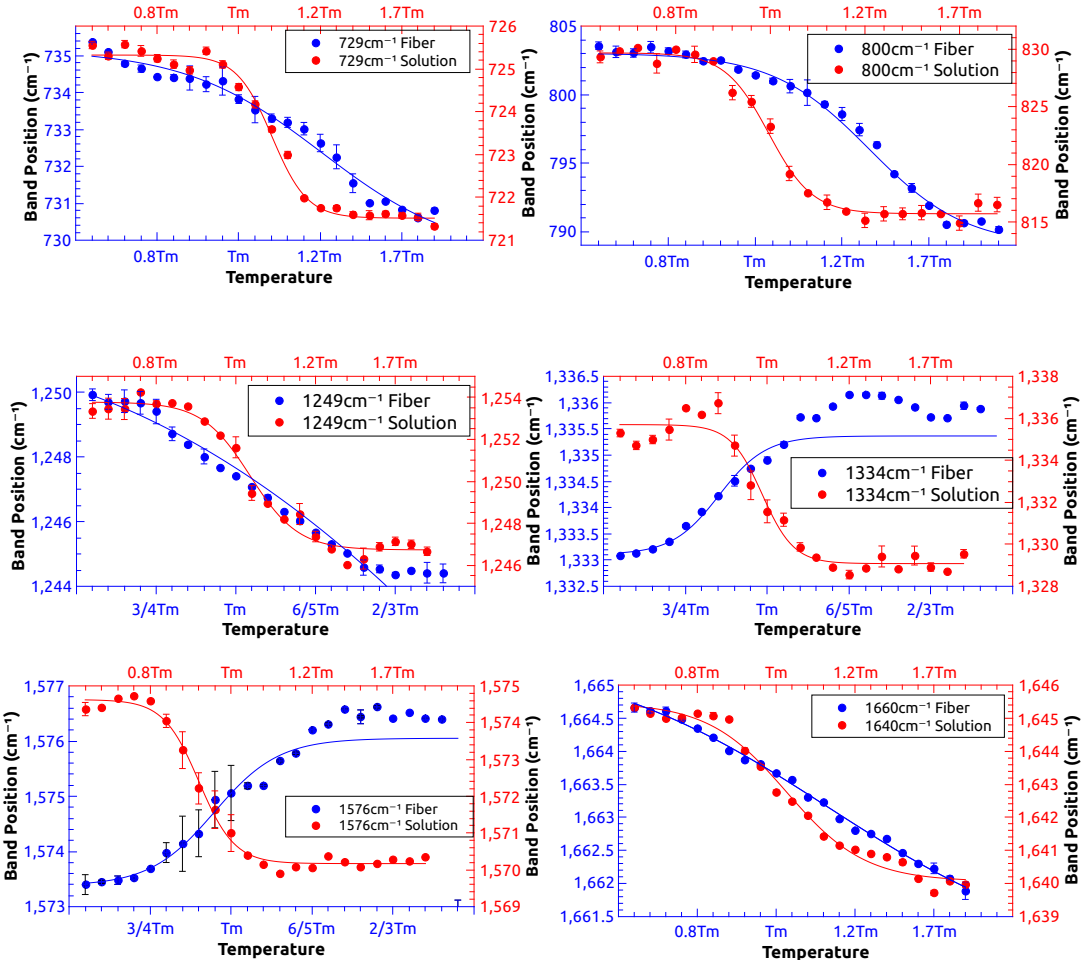


Fig. 5.20 Temperature-dependent relative Raman frequency shift of DNA dry fiber (blue) and DNA solution (red) of the $729, 803, 1249, 1334, 1576$ and $1660 cm^{-1}$ vibration modes of DNA. All frequencies result shown, represent an average which correspond to three independent experimental realization. Error bars are shown for each data.

Table 5.4 Width of the melting transition for DNa fiber and DNA solution obtained by fitting denaturation curves of characteristic selected bands.

Type of band	**Position** (cm^{-1})	**Width Fiber**	**Width Solution**
Base pair	729	5.36	3.08
	1660	6.3	5.84
Sugar ring	1335	4.26	2.48
	1570	5.27	2.19
Backbone	800	7.44	2.229
	1248	4.12	3.89

On sight of results listed on Table 5.4 The average width of the transition is 5.45 ± 1.47 for DNA fiber and 3.28 ± 1.58 for DNA solution. As we mentioned before, this results suggest a smooth and continuous transition in the fiber, as reported previously [20, 107] and really close to the data obtained in the A/B-form fiber in the previous section. However, for DNA solution transition seems to be sharper and occurring narrow range of temperature, according with the theory of DNA being a one dimensional system [11].

This section has been dedicated to the comparative study of the melting transition of DNA fiber and DNA solution by Raman spectroscopy. For this study we can conclude that there is some specific bands in DNA Raman spectrum that allow to analyze melting transition and help us to understand how it works the denaturation process in both states, fiber and solution. By analyzing the frequency shifts occurred in this determined bands, we have been able to compare structural changes produced in each state and to study the differences in the melting transition between them. The differences between both states are mostly related with bands assigned to backbone conformation which is explained because structural restrictions in fiber and more freedom or flexibility in solution, and to the elimination or formation of new hydrogen bonds that are easier in solution. We also have shown that the calculated width along the melting transition takes part, is in excellent agreement with previous works (even our own results in previous section) and underlines the differences between DNA fiber and DNA solution.

5.5 Raman analysis of the solvent influence in DNA fibers.

Until now, we have studied melting transition of DNA dry fibers and DNA solution paying attention to the resonance information of DNA base pairs by Raman spectroscopy. We also have compared how the transition takes part in both states, fiber and solution correlating structural changes occurred in each state. Our aim now, is to study the solvent influence in DNA fibers since our work team has studied the melting transition of this fibers submerged in different solvents by neutron scattering, besides we think it could be an interesting information itself.

Based on neutron experiments we are going to study the influence of different solvents such as water (relative humidity) or poly(ethylene glycol) (PEG) whose molecules can exerts osmotic pressure on the fibers [128, 129] producing some confinement in Na-DNA fibers and affecting DNA intermolecular distance and thus, melting transition.

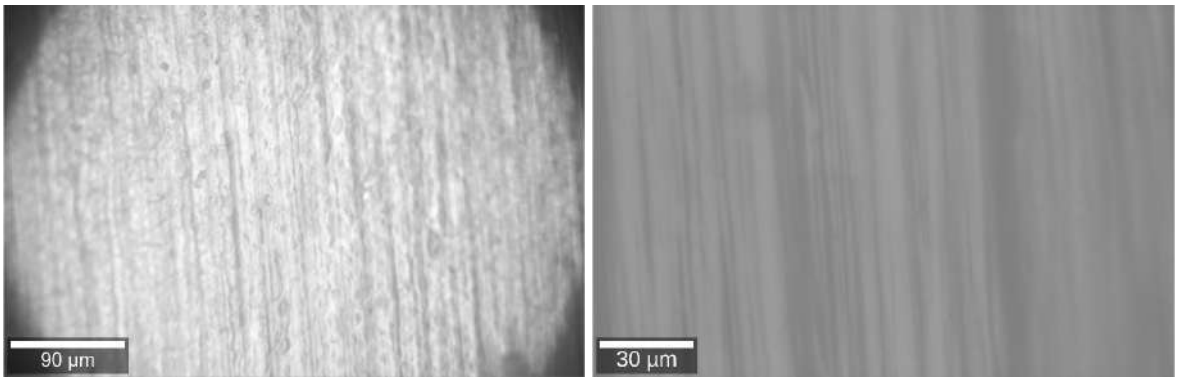
The studied that we have carried out is to compare DNA dry fiber with a humidified fiber and fibers submerged in different PEG solutions, by analyzing the frequency shifts of

the Raman bands present in their spectrums at room temperature. Dry fiber, as in previous sections, was obtained with the wet spinning method [74] and stored in a supersaturated NaCl solution environment which provides a 75% relative humidity atmosphere. This is a re-humidification processes for obtained the final DNA films setting the water content and stabilizing its structure. The second set of samples, the humidified fibers, was obtained taking dry fibers like starting material and humidifying it to 92% in a H_2O atmosphere. This atmosphere ensures that DNA molecules adopt the B-conformation [119]. Finally, last set of samples consisted in the same DNA dry fiber submerged in 6000MW PEG solution with different concentrations. PEG polymer was dissolved in a buffer consisting in 0.1M NaCl, 10mM Tris and 1mM EDTA and we uses concentrations of 17,25 and 40% w/w.

5.5.1 DNA dry fiber vs DNA humidified fiber

Firstly, we compare DNA dry fiber with humidified fiber. Beforehand these two types of fibers are differentiated by the final water content with the dry fiber being a mixture of DNA forms (mostly A/B) and the wet fiber only B-form [119]. We will see some differences in the frequency shift of Raman resonances of both fibers.

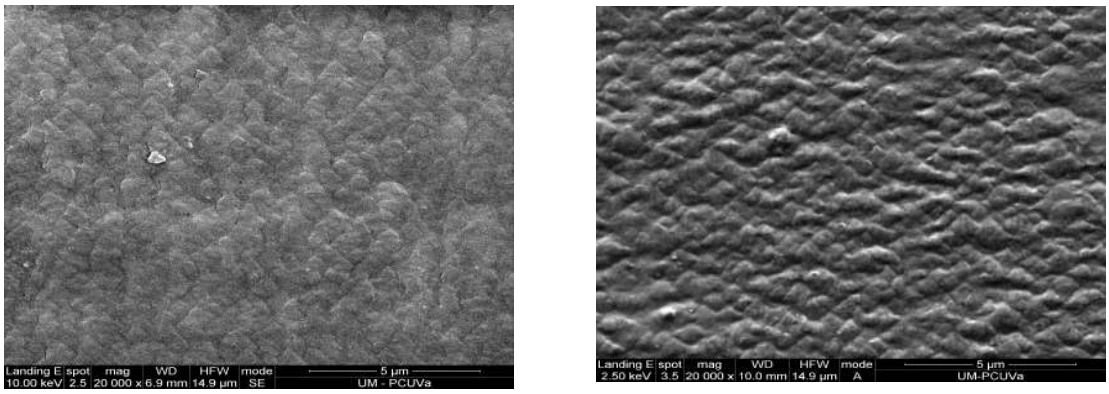
For the Raman spectrums, the samples were introduced in a sealed glass capillary the measurements was recorded with an integration time of 30sec, 10 accumulations and laser power of 20mW. We use a confocal micro-RAMAN *Alpha* 300 *R* instrument (Witec) explained in section 5.1.1 which has incorporated an *Olympus* optical microscope, so we did a first physical comparative analysis with the images provides by the microscope. This pictures are depicted in Fig. 5.21



(a) 20X objective lens (b) 50X objective lens

Fig. 5.21 Microscope view of fibers obtained with Olympus optical microscope. Left panel shows image for DNA dry fiber while rigth panel presents humidified fiber picture.

Both images shows the general morphology of the fibers seeing the orientation of it, but DNA humidified fiber soak up some water and it is more expanded. It remember to a thermolabile deformation suffered by a plastic material. Similar results can be observed in Fig. 5.22 which shows an extra characterization of the fibers done by SEM microscopy. In Fig. 5.22 we can observed SEM images where we can appreciate better the texture of the fiber, and what could be some kind of agglomerations consisting on many DNA molecules together in the same orientation. As we said above, the picture belonging to humidified fiber, exhibits a more swollen texture. The texture is completely different compare with dry fiber due to the absorption of water.



(a) Dry fiber (b) Humidified fiber

Fig. 5.22 Scanning Electron Microscopy (SEM) pictures of DNA fibers using the beam deceleration mode (BMD) with a Microscope $(ESEM)$ $FEI-Quanta$ $200FEG$. Again left panel shows DNA dry fiber and rigth panel corresponds to DNA humidified fiber.

After the previous physical characterization that leading us to make a first impression of what to wait with these different fibers, we measure Raman spectrums of each sample at room temperature. The result is shown in Fig. 5.23. At this point, we recognize perfectly the main bands present in the Raman spectrum, which we can observe that appear in the Fig. 5.23.

Due to each fiber gives different signal, first of all we normalize the spectrum in order to be comparable. We can observe that the bands are similar in both cases but some of them suffer a frequency shift as it happened in solution. Again for easier localization we listed the shifth reflected in Fig. 5.24 in the following table.

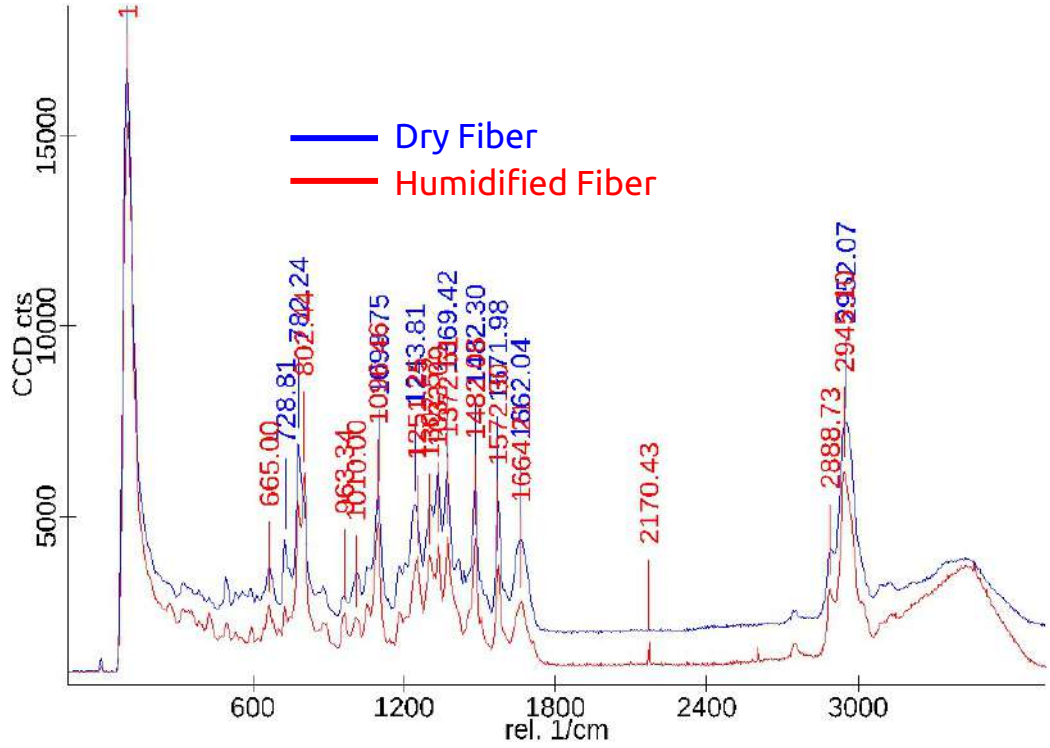


Fig. 5.23 Spectrum of DNA Dry Fiber (blue) and DNA humidified (red) recorded at room temperature showing main bands.

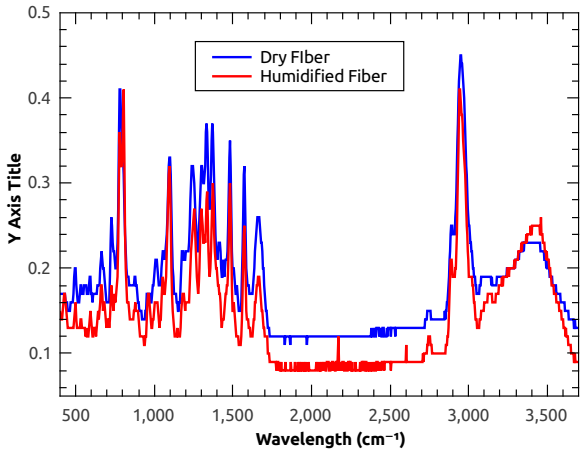


Fig. 5.24 Spectrum of DNA Dry Fiber (blue) and DNA humidified (red) recorded at room temperature normalized in order to intensities be comparable.

Table 5.5 Wavelength (cm^{-1}) of the peaks of Raman spectrum of Fig. 5.24 for DNA dry fiber and DNA humidified fiber. The quantified frequency shift is shown.

Dry Fiber	**Humidified**	$\Delta\sigma^2$ **Dry**	$\Delta\sigma^2$ **Humidified**
666	665	0.396	1
782	779	1.443	3
1098	1096	0.44	2
1243	1251	0.47	8
1331	1334	0.74	2
1482	1482	0.43	0
1571	1572	2.29	1
1662	1664	1.79	2

Again, we define new parameters that help us to identify the most relevant shifts occurred in the spectrum. This variables are also shown on Table 5.5.

$$\Delta\sigma^2_{shift} = |Band\ Humidified fiber\ RT - \ Band\ Dry fiber\ RT| \tag{5.8}$$

This parameter, equal to the one defined by Eq. 5.7 give us an idea of the structural differences between dry fiber and humidified fiber, comparing the bands present in Raman spectrum at room temperature of them. $\Delta\sigma^2$ Dry is the reference value of the dry fiber obtained with different measured spectrums of the dry fiber at the same temperature.

Looking at the results of Table 5.5 that summarizes results of spectra shown in Fig. 5.24, it can be observed the bands that experiment the bigger changes related with the frequency shift at room temperature between dry and humidified fiber. These bands are $782, 1098, 1243$ and $1331 cm^{-1}$. This differences are also appreciables in Fig. 5.25

Band at $782 cm^{-1}$ corresponds to phosphodiester marker and for residues of dC being a diagnostic for base stacking in cytosine[113, 127]. The difference between the two kind of fibers is not really big, but in the case of the humidified fiber where it seems that the molecule absorbs some water doing the chains having less spacial confinement, making cytosine residues more accessible. In dry fiber both dC and the phosphate backbone are more restricted so the difference between both fibers is evident. Vibration mode at $1100 cm^{-1}$ is assigned to the symmetric phosphodioxy (PO_2) stretching vibration. Thus, the reason of the difference present in both fibers could be the same and it is related with structural

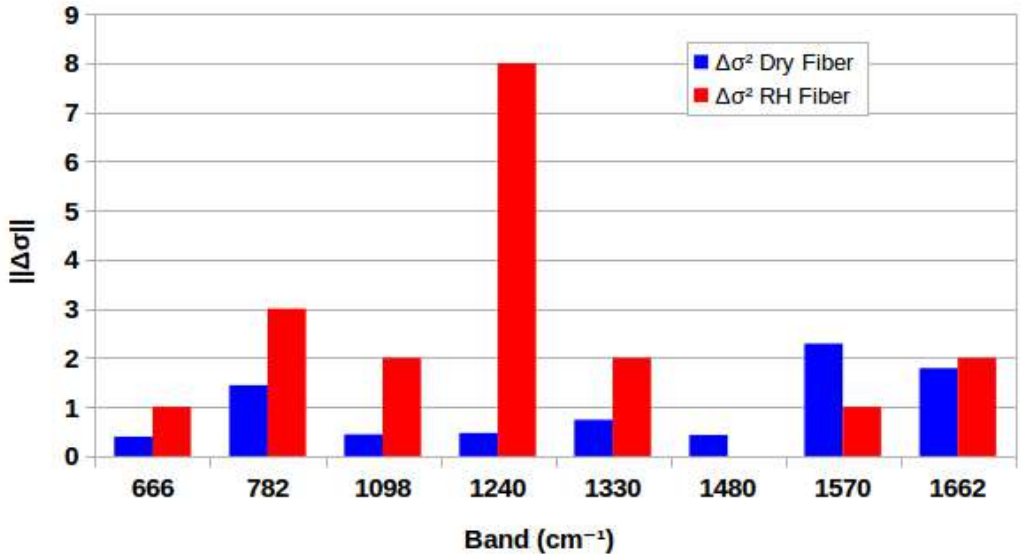


Fig. 5.25 Differences in the parameters defined by Eq. 5.7 at room temperature. This parameter represent the shift suffered by the main bands present in DNA Raman spectrum of both studied fibers.

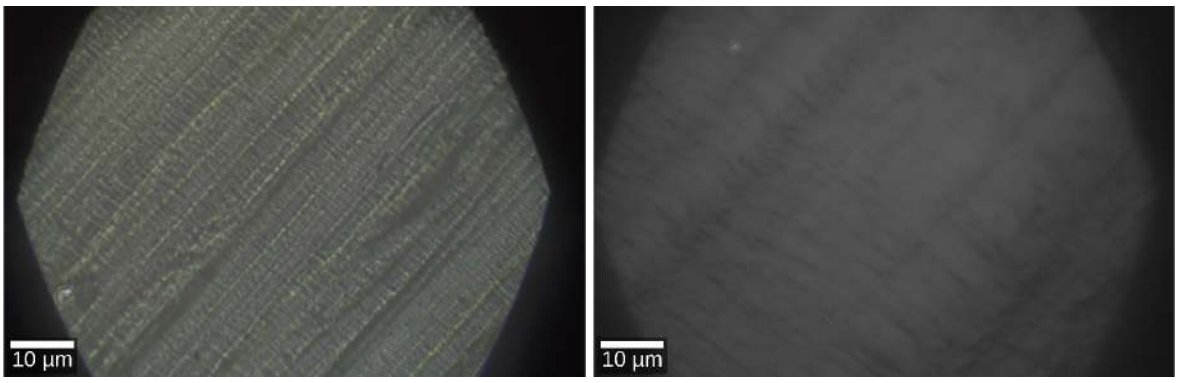
restrictions. PO_2 groups are sensitive to interactions of metal ions and solvent molecules so in the humidified fiber where the interaction with the solvent is bigger, this band suffer a appreciable shift. Regarding to band at $1249cm^{-1}$, this is a mixture of the adenine exocyclic C6-N6 bond and the N1=C6 ring bond and a marker of dT unstacking[110, 111, 115].This band is also a marker of B-form [112, 114] therefore it is once again more affected in DNA humidified fiber since is completely B form while dry fiber is a mixture. Finally, band situated at $1331cm^{-1}$ is assigned with stretching vibrations of imidazole ring [120] thus related with canges in dA and dG residues. The shift that happen in this band is related with the formation of new hydrogen bonds (or the elimination of other already formed) so makes sense that in the humidified fiber this residues could form more hydrogen bonds.

Summarizing, we can conclude that the main changes in the frequency of the Raman bands at room temperature between dry and humidified fiber is due to the molecular confinement and interaction with the solvent. Humidified fibers where DNA molecules adopt B-form, present bands related with markers of this forms while dry fiber present the same bands with a shift that could indicate A-form. In addition, humidified fiber present less structural restriction, DNA molecules are more expanded since they have more spatial freedom, and can have more interaction with solvent resulting in the formation of new hydrogen bonds.

5.5.2 DNA dry fiber vs DNA fiber submerged in PEG solution

Now, we compare DNA dry fiber with the same fiber submerged in a PEG solution with different concentrations. This will allow us to study the effect of the osmotic pressure and the spatial confinement exerted on the DNA molecules. We compared the dry fiber with three different samples of fiber submerged in a 17,25 and 40% w/w PEG solution.

Like we did before with the humidified fiber, we first compare the different samples with optical microscopy and SEM. The comparison of microscope view it can be observed in Fig. 5.26.



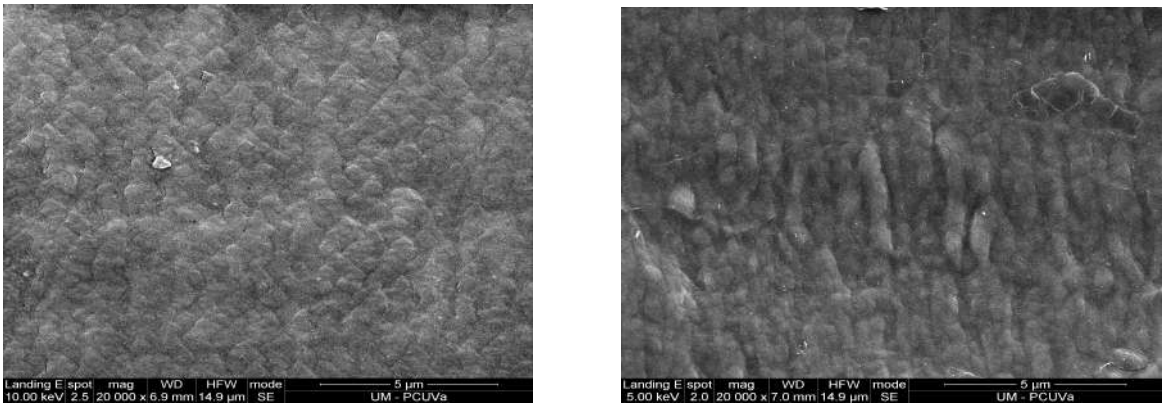
(a) 20X objective lens (b) 100X objective lens

Fig. 5.26 Microscope view of fibers obtained with Olympus optical microscope. Left panel shows image for DNA dry fiber while right panel presents fiber submerged in PEG picture.

Looking at right panel, we can see that PEG varies the intermolecular distance of DNA fibers. The PEG presence stabilizes the molecular conformation since occupies the space around helix molecule [130, 131]. In this range of concentrations PEG not really penetrate the fibers being only outside in contact with DNA molecules in the surface of the sample. But in the presence of PEG the interaxial distance in the fibers changes [128] as it is reflected in Fig. 5.26. In addition, despite in both images it is observed the orientation of the fiber, it seems that fiber submerged in PEG solution exhibits a softer texture.

Images obtained with SEM reflect similar results. To preparing the samples to recorder the images, we fix the fiber to a piece of mica with special tape and then we submerged it in a solution of 17%PEG. Also we metalized the sample with a thin layer of gold. For this samples, we use high vacuum with secondary electrons method.

In the PEG image it can be observed clearly the filaments which form the fiber and could be because the use of PEG solutions allowed the fibers to swell without losing their orientation. As we said above, PEG solution changes the confinement of DNA molecules by osmotic pressure being able to see individual filaments. But again, we can perceive



(a) Dry fiber (b) PEG fiber

Fig. 5.27 Scanning Electron Microscopy (SEM) pictures of DNA dry fiber using the beam deceleration mode (BMD) (left panel) and DNA fiber submerged in 17% w/w PEG solution using secondary electron method (right panel).

agglomerations of many molecules in each filament although keeping the orientation of the fiber.

With the different samples again introduced in a sealed glass capillary, we recorded Raman spectrum of each sample with the same conditions, integration time of 30sec, 10 accumulations and laser power of 20mW. Again, first of all, we normalize the spectrum in order to be comparable and the final result is plotted in Fig. 5.28.

The first thing that attracts our attention in Fig. 5.28 is that in the Raman spectra we also detect the characteristic frequencies of the vibrations of the PEG molecules. These will serve as a filter, or as selectable bands to try to elucidate the distribution of the polymer with respect to fiber.

We can observe that the bands are similar in all cases but some of them suffer a shift, normally to higher wavelengths, being the bigger shift in the fiber with 17% of PEG like we can see reflected on the Table 5.6. It can be also observable that all concentrations present the same bands in the same positions, suggesting that the concentration of PEG does not affect the way that the polymer interacts with the DNA.

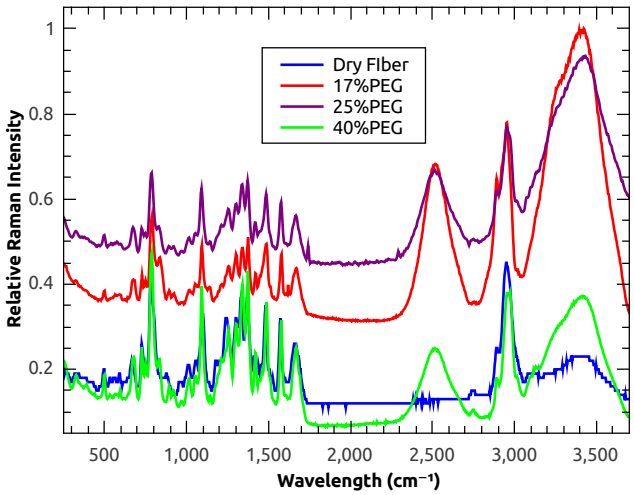


Fig. 5.28 Raman spectrum of DNA dry Fiber (blue) and fiber submerged in different PEG solutions, 17% (red), 25% (purple) and 40% (green) recorded at room temperature. The spectrums were normalized in order to intensities be comparable.

Table 5.6 Wavelength (cm^{-1}) of the peaks of Raman spectrum of Fig. 5.28 for DNA dry fiber and fiber submerged in different PEG solutions, 17,25 y 40% w/w. The quantified frequency shift is also shown.

Dry Fiber	**17% PEG**	**25% PEG**	**40% PEG**	$\Delta\sigma^2$**17%**	$\Delta\sigma^2$ **25%**	$\Delta\sigma^2$ **40%**
666.54	678.28	673.27	678.66	11.74	6.73	12.12
782.24	789.38	785.62	786.64	7.14	3.37	4.40
1098.75	1094.46	1091.13	1092.68	4.29	7.62	6.07
1243.81	1254.91	1253.8	1254.93	11.1	9.99	11.12
1331.48	1341.08	1338.43	1339.98	9.6	6.95	8.5
1369.42	1373.28	1372.91	1374.05	3.86	3.49	4.63
1482.3	1487.31	1484.37	1487.77	5.01	2.07	5.47
1571.98	1577.96	1575.32	1576.87	5.98	3.34	4.89
1662.04	1668.14	1665.52	1669.45	6.1	3.48	7.41

As we did with humidified fiber and also with DNA solution, we define a new variable that allow us to quantify the shift of the bands and correlate these changes with structural differences between dry fiber and fiber submerged in PEG solution. This parameter, listed on Table 5.6 is defined by Eq. 5.8

$$\Delta\sigma^2_{shift} = |Band\ PEG\ RT - \ Band\ Dry fiber\ RT| \tag{5.9}$$

According to the results of parameters defined by Eq. 5.9 DNA submerged in PEG buffer suffer a shift in the bands to higher wavelengths (only the band at $1098cm^{-1}$ moves to lower lengths), and with different increment according to PEG percentage, being the least important change usually the one that corresponds to 25%. This can be seen better in the Fig. 5.29

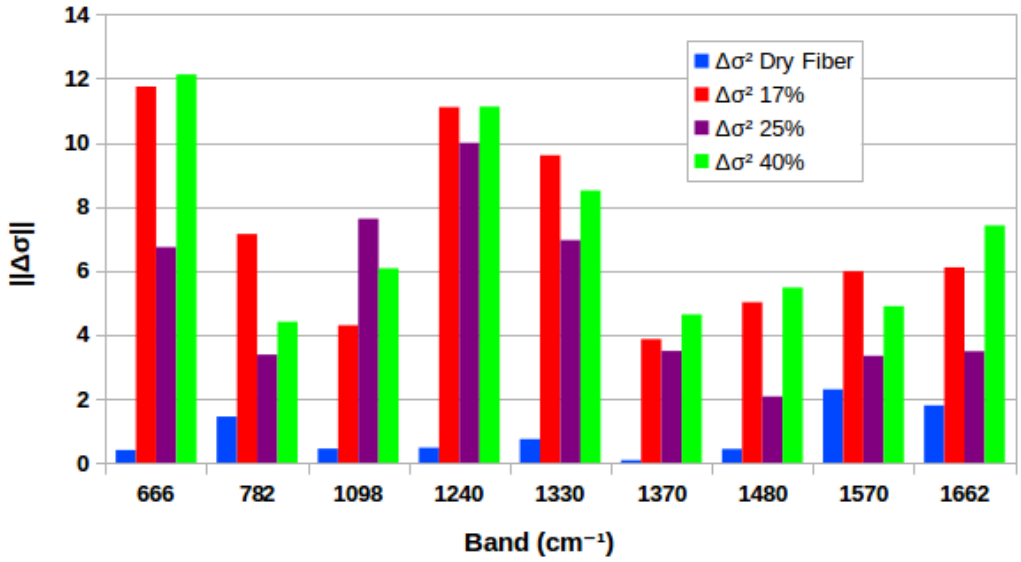
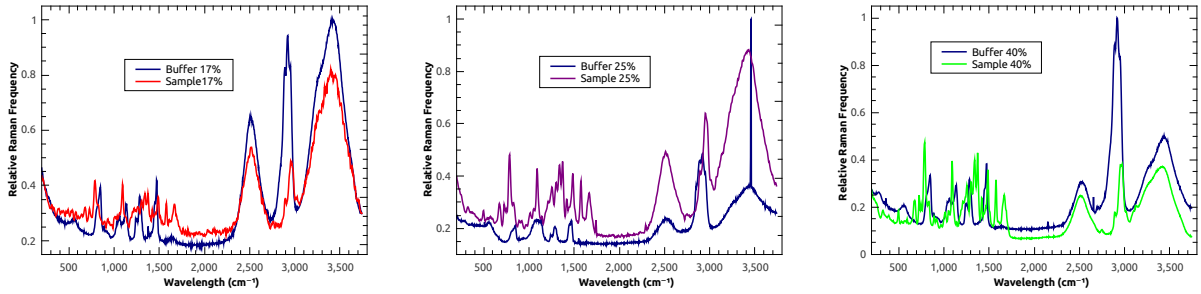


Fig. 5.29 Parameters defined by Eq. 5.9 at room temperature, that quantify the shift suffered by the main bands present in DNA Raman spectrum of dry fiber and fiber in PEG solution.

All the main bands exhibit a pronounced shift. Let's explain why could be each one. Band at $666cm^{-1}$ is assigned to a dG residues marker [110, 112, 113]. This band suffer a shift to higher wavelengths contrary to what happen in DNA solution or DNA humidified fiber. Due to the osmotic pressure that exerts PEG molecules could happen that guanine residues are less accesibles. $780cm^{-1}$ band is assigned to backbone $O-P-O$ of A-form. In solution and humidified fiber this band moves to lower wavelengths but in PEG solution the shift is to upper. Again, fibers submerged in PEG present different structural restrictions since the presence of PEG increase the intermolecular distance with respect to dry fiber or humidified one. The confinement of the DNA molecules are thus different in submerged fibers and this is reflected by a change in this band frequency. This effect changes with PEG percentage, where higher PEG concentration decrease the intermolecular distance

[128]. It is because that, we see in Fig that the shift for $780cm^{-1}$ band is higher for 17% of PEG. This fact is going to explain most of the localized shifts in Raman frequencies. Band at $1098cm^{-1}$ is assigned to the symmetric phosphodioxy (PO_2) stretching vibration. Here, the shift occurred in submerged fibers is to lower frequencies equal to what happened in humidified fibers. Thus, the reason could be the same and it is related with structural restrictions. We now know that PEG makes higher the intermolecular distance so PO_2 groups, which are sensitive to interactions with solvent molecules, could have bigger interaction with solvent, and this band suffer a appreciable shift. In this case the biggest change corresponds to a concentration of 25% but it is not the usual. Vibrational mode at $1249cm^{-1}$ is a mixture of dT unstacking marker[110, 111, 115] and B-form marker[112, 114]. As it happened in solution and humidified fiber, this bans suffer a shift to higher wavelengths respect to dry fiber frequency. This is due one more time because an increment in the intermolecular distance allows thymine residues be more exposed. The same behavior is shown for band at $1331cm^{-1}$ corresponding with stretching vibrations of imidazole ring [120], that is, adenine and guanine residues. In addition to the increment of the intermolecular distance, this band is related with the formation of new hydrogen bonding that could happen easier in fibers submerged. Bands at $1369, 1482$ and $1571cm^{-1}$ corresponds respectively to very remarkable C4=O of dT [111, 112], N7 acceptor of guanine ring [110, 111, 115] and ring mode of dA and dG [111–113, 115]. All of them are related with elimination or formation of hydrogen bonds and present the same behavior. In dry fiber DNA molecules present a diameter of 20^{-1} while it has been reported that in the presence of PEG the immersed fibers present distances between the neighboring molecules of a few more Angstroms[125]. Therefore, the original hydrogen bonds could change making this bands suffer a moderate shift. Finally, the last shift recorded at room temperature occurs in the broad band centered $1662cm^{-1}$. This band is attributed to the carbonyl group of the nucleotides dG, dC and dT, the latter being the one that contributes the most[111, 113, 118]. It is known that this band is very sensitive to solvent interactions and the disruption of hydrogen bonding. Thus, the shift to higher frequencies of this band, that we see in submerged fibers, could reflect different contributions of thymine carbonyl acceptor when it is bonded with water molecules or adenine residues [115, 126].

In order to study possibles differences with different concentration of the PEG solution and how this interacts with DNA fiber, for each concentration, we plot the spectrum of the fiber submerged in PEG with its corresponds spectrum of only buffer. Before that, we normalize both because the intensity of each one is really different and we cannot compare it. This spectrums are shown in Fig. 5.30.



(a) 17% PEG (b) 25% PEG (c) 40% PEG

Fig. 5.30 Raman spectrum of fibers immersed in PEG solution together with the spectrum of only correspondent buffer

We can see clearly which bands corresponds to the PEG (or water) and which one to the DNA itself, so it is useful for following analysis. A priori we do not see big differences between each PEG concentration beyond the intensity of some bands and this could be due for the set up of the instrument.

Next image show again a Raman spectrum of a fiber submerged in PEG solution, but in this case we select different peaks, ones corresponding to classical DNA bands and others that only appear in the buffer. With this selected peaks, we create a 'mask' that allow us to obtain a 2D image trying to identify how the polymer interacts with the DNA.

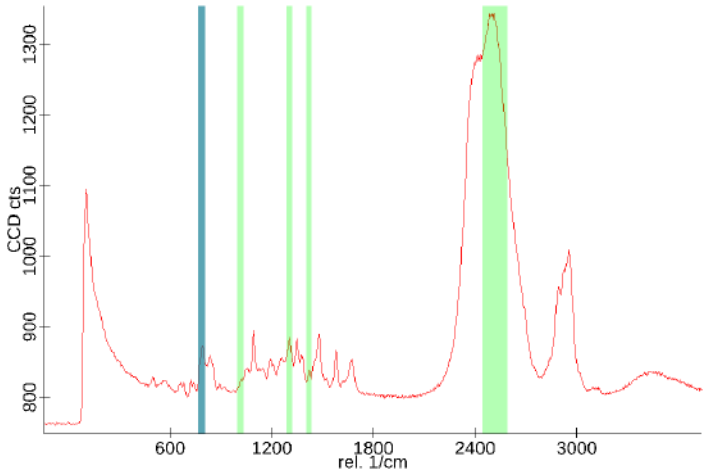


Fig. 5.31 Fiber immersed in PEG solution Raman spectrum with differents bands of DNA and PEG selected.

The result of the selected filters with characteristic bands of both, DNA and PEG buffer, is depicted in Fig. 5.32. In this 2D images it can be observed the distribution of the PEG and the DNA into the fiber. First row corresponds to filters selected with main bands present in

PEG (and that not appear at all in DNA spectrum). Contrary, second row exhibit the selected bands of DNA that has not analogy with PEG peaks. Finally, last row is the combination of both filters, DNA and PEG for each sample, i.e DNA submerged in PEG solution of different concentration. Paying attention to the bar scale of each picture, we can sense the distribution of PEG and DNA respectively. It is more clear in filters $2516cm^{-1}$ for PEG and $786cm^{-1}$ for DNA in whose images we can observe some vertical orientation that indicates the distribution of the PEG on the fiber. This distribution can be shown by darkness zones.

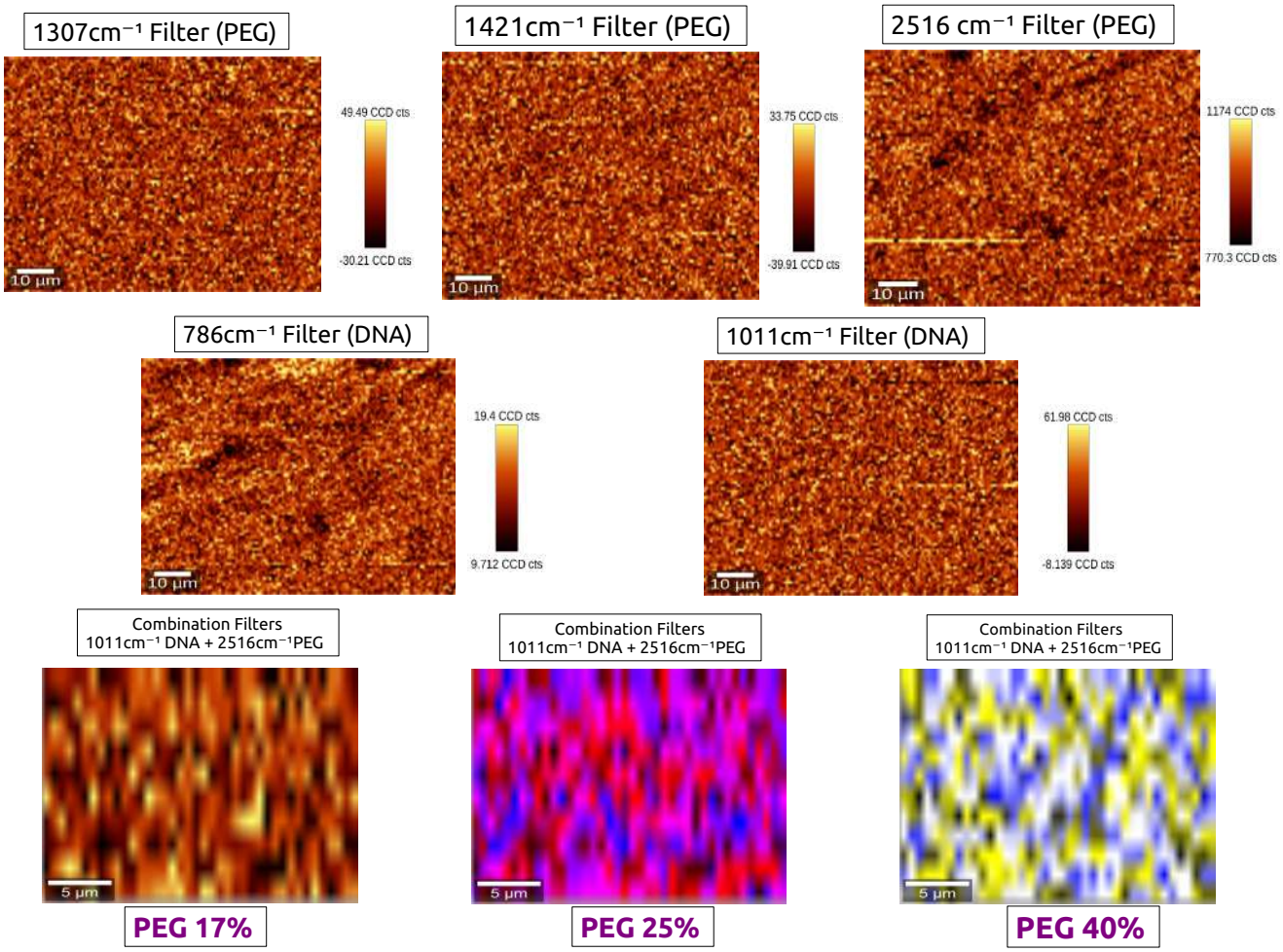


Fig. 5.32 2D images of distribution of PEG and DNA in fiber samples. First row is for PEG selected filters, second row for DNA bands and last row the combination of two of them for each sample.

The combination of the filters shown in last row of Fig. 5.32 for each sample, presents the same results. A homogeneous distribution of the PEG on the fiber can be discerned, but maintaining the slight diagonal orientation discussed above.

The conclusion that can be extracted for this study is that fibers submerged in PEG solution stabilize, the filaments that forms the fiber are more separated and the intermolecular distance between molecules increase. In addition, the distribution of the PEG solution is homogeneous, at least over sample surface (at this concentrations PEG do not really penetrate into fibers free spaces) and there is some characteristic bands that allow us to identify this structural changes and may help to see PEG distribution. Finally, we have seen that most of the produces changes are bigger with less concentration of PEG (17%) since the increment of the intermolecular distance is higher for less concentration.

5.5.3 DNA dry fiber vs DNA fiber submerged in ethanol solution

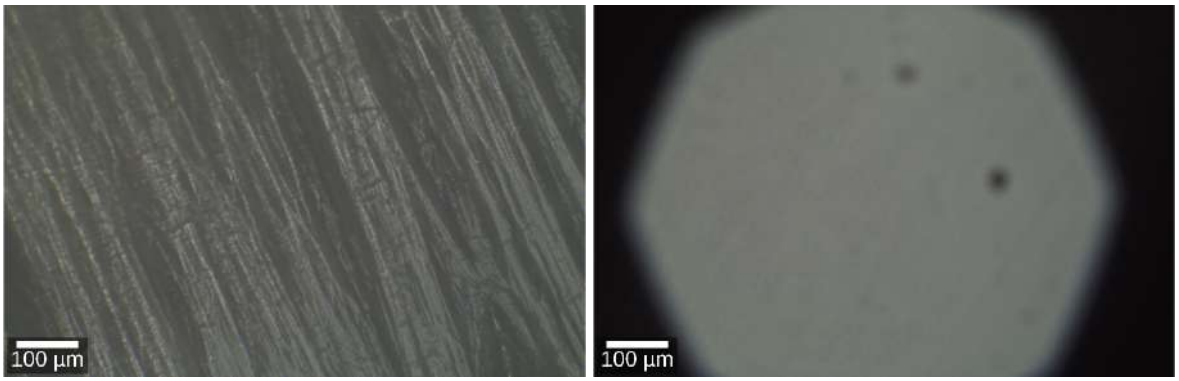
In addition to the PEG study, we also want to analyze the effect in the dry fiber when is submerged in ethanol (EtOH) solutions. EtOH molecules also can exert an osmotic pressure over the fiber modifying the confinement of the molecules. Moreover, ethanol increases the electrostatic repulsions among phosphate groups of the helix due to the lowering of the dielectric constants of the ethanol solution. Our work team has done some studies of the melting transition of DNA fibers immersed in ethanol solutions by neutron scattering.

Although the main idea is to test several different concentrations of ethanol as we did with PEG, we made a first study at room temperature comparing the dry fiber with a 60% ethanol concentration.

One more time, we characterize both samples, dry fiber and fiber immersed in ethanol, in order to appreciate some physical characteristics that allow us to correlate structural differences that will be reflected in differences in Raman spectrum. Fig. 5.33 shows the results of physical characterization by optical microscopy and Fig. 5.34 for SEM pictures.

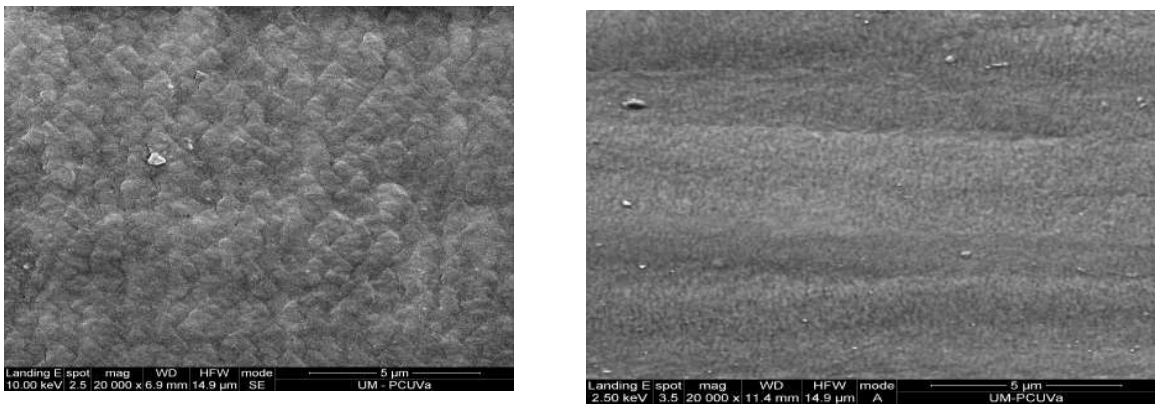
In the microscope view, fiber immersed in ethanol is not really well appreciated. But in the SEM picture we can see that with the ethanol, the fiber is expanded. It is swollen and it can be observed more fibrillate appearance. In both, dry and in EtOH solution, we observe some orientation but in the ethanol sample there is some zones with different texture which could be due to the ethanol is not distributed homogeneously.

With ethanol the fibers separate because they swell but less than with PEG and in a different way. Unlike PEG, ethanol destabilizes DNA, which causes the bands of the Raman spectrum to undergo a shift just as it did with PEG because it changes the intermolecular distance and the confinement of the molecules, but in a different way. In addition high concentrations of ethanol cause DNA dehydration causing a mixture of A / B forms. Ethanol is a very small molecule that will penetrate the fibers as easily as the water molecules.



(a) Dry fiber (b) Fiber submerged in ethanol solution

Fig. 5.33 Microscope view of fibers at low magnifications. Left panel shows image for DNA dry fiber while right panel presents fiber submerged in ethanol picture.



(a) Dry fiber (b) Fiber submerged in ethanol solution

Fig. 5.34 SEM images for DNA dry fiber (left) and DNA fiber submerged in ethanol solution of 60% (right).

Next to do is recorder Raman spectrums in order to analyze frequency shift produced in the bands between dry fiber and fiber in ethanol. The measurements were at room temperature in a glass capillary sealed. We compare DNA dry fiber with fiber submerged in a 60% ethanol solution. We also analyze differences between fiber in ethanol with only its buffer in order to localize specifics bands of sample and buffer to, as we did with PEG, create 2D images to study ethanol distribution. The results are plotted in Fig. 5.35.

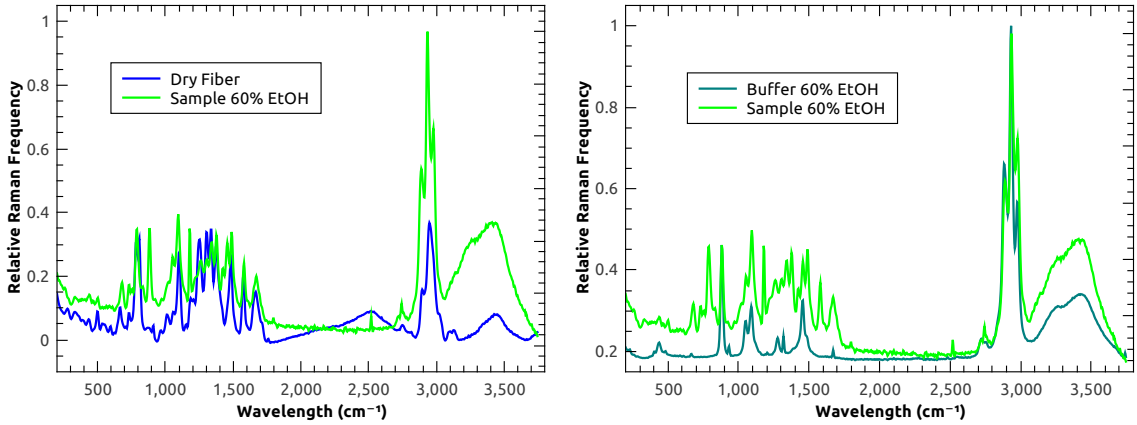


Fig. 5.35 Left panel shows Raman spectrum of DNA dry Fiber (blue) and fiber submerged in a 60% ethanol solution (green). The spectrums were normalized in order to intensities be comparable. Right panel is the Raman spectrum of 60% ethanol fiber sample with its buffer.

In the left panel of Fig. 5.35 we can observe Raman spectrum of DNA dry fiber (blue line) and the sample consisting in a DNA fiber immersed in 60% of ethanol solution (green line). First thing observable, is once again, the shift present in some bands between dry fiber and ethanol sample. For easier localization, we list this frequencies for both samples on Table. 5.7 and again we quantify the shift occurred with the parameter defined in Eq. 5.10. Right panel shows ethanol sample with its corresponds spectrum of only ethanol buffer which then will allow us to select different bands in order to plot 2D images.

$$\Delta\sigma^2_{shift} = |Band\ EtOH\ RT - Band\ Dryfiber\ RT| \tag{5.10}$$

Table 5.7 Wavelength (cm^{-1}) of the peaks of Raman spectrum of Fig. 5.35 for DNA dry fiber and DNA fiber in 60% ethanol solution. The quantified frequency shift is shown.

Dry Fiber	**Ethanol 60%**	$\Delta\sigma^2$ **Dry**	$\Delta\sigma^2$ **EtOH**
784.52	787.38	1.443	2.841
805.85	883.27	1.2	77.42
1101.37	1094.41	0.44	6.69
1250.83	1260.98	0.47	10.15
1337.61	1341.11	0.74	3.5
1485.12	1489.48	0.43	4.36
1576.06	1580.01	2.29	3.95
1665.21	1671.04	1.79	5.83

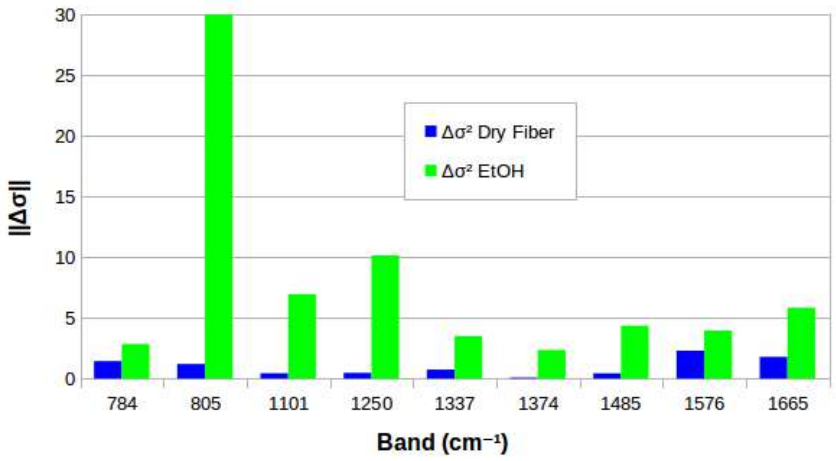


Fig. 5.36 Parameter defined by Eq. 5.10 at room temperature, that quantify the shift suffered by the main bands present in DNA Raman spectrum of dry fiber and fiber immersed in 60% ethanol solution.

We can notice that all the bands of fiber submerged in ethanol solution suffer a shift to higher wavelengths respect to DNA dry fiber. As we mentioned above, with ethanol, fibers swollen and thus are more separate (there is an increment of the electrostatic repulsions and thus a change of the intermolecular distance). Namely, although the underlying mechanism and other important features are different, the shift occurred in Raman peaks can be explain for the same reasons that we mentioned with PEG. Therefore, we are not going to explain again each band. Only it is necessary to keep in mind that the changes in the bands are related with structural changes (backbone and DNA form markers are affected) and with

the elimination or formation of different hydrogen bonding. This fact is again because the change in the intermolecular distance, the hydration of the molecules that form the fiber (and the conformation that they adopt) and the confinement of the molecule.

In order to discern if we can learn more about the distribution of the ethanol solution in the sample, we create new 2D images with filters selected from both the buffer and the fiber as we did with the PEG. Again, we select peaks of ethanol that do not appear in DNA and vice versa, characteristic bands of the biological molecule that are no vibration modes in the buffer.

2D images consisting in different selected filters of the samples are shown in Fig. 5.37. First row corresponds to the filters applied with ethanol characteristic bands, second row to DNA main bands and third row the combination of both in the 60% ethanol sample. For ethanol buffer we select bands at 880 and $1450 cm^{-1}$ while for DNA the peaks are at 1377 and $1576 cm^{-1}$.

In the images corresponding to the ethanol filters, we can see a distribution in half if we divide the image diagonally. This could be a signal of the distribution of the ethanol is not really homogeneous. Contrary, in the images of DNA bands it is observable some kind of horizontal orientation suggesting the fiber axis in this direction. In the third row color blue correspond to DNA and is distributed again horizontally, while red zones are related with ethanol and now seems more homogeneous despite there is some zone with higher tonality.

Therefore, for fibers submerged in ethanol solution we can conclude that DNA is destabilized causing the swelling of the fiber and increasing the intermolecular distance between DNA helices, which implies a shift of the frequencies of the Raman bands to a higher wavelengths. In addition ethanol causes a confinement and a dehydration of the DNA molecules, but the distribution in the fiber can be homogeneous. More studies, and with different ethanol concentrations, are necessary to draw better conclusions and compare them with the study of neutrons.

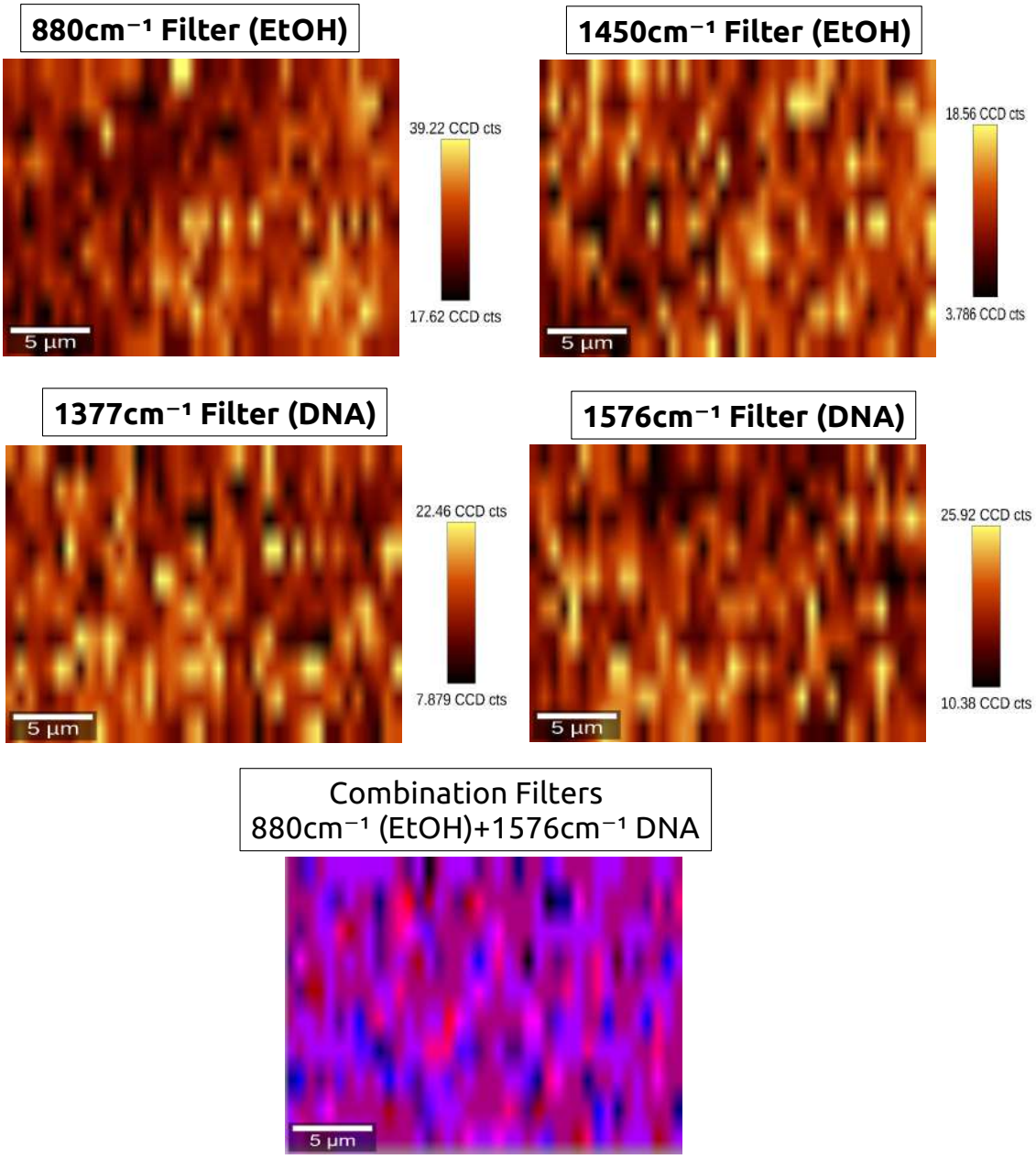


Fig. 5.37 2D images of distribution of ethanol and DNA in fiber samples. First row is for EtOH selected filters, second row for DNA peaks and last row the combination of two of them for the sample.

Chapter 6

Conclusions

6.1 Conclusions

The thermal denaturation of DNA is a fundamental process not only because of its biological importance, since it initiates vital processes for life, but also involves a phase transition in a dimension being a very interesting process itself from a physical point of view. DNA is a very dynamic molecule, which can undergo fluctuations even at biological temperatures. These fluctuations can lead to the formation of denaturation bubbles, and if the temperature continues to rise, reach the complete separation of the strands that form the DNA helix, in a process called the DNA melting transition. This phenomenon is the main topic addressed during most of this doctoral thesis. The aim of this work has been to obtain information about thermal denaturation transition, be it theoretical data about opening constants or structural information during the transition.

In Chapter 3, a theoretical study on dynamic properties of DNA has been carried out, such as denaturation using the PBD model [11] described previously in Chapter 2. Based on previous studies conducted by the group, we have added an entropic barrier to the Morse potential of the original PBD model, which takes into account the dependence of the stacking of the dimer base pairs [19]. This modification of the model allows a quantitative improvement in the opening rates of DNA hairpins.

The model implemented has been used to calculate the opening constant of a special DNA architecture such as DNA hairpins. This structure has been chosen because of its importance, given its involvement in biological processes, as well as in biotechnological applications. DNA beacons are being used as nanodevices, biosensors or optical markers to detect genes or markers and inhibitors of sequences, and as molecular memories or RAM memories.

The complete DNA denaturation is difficult to obtain using standard molecular dynamics (MD), since it is a rare event in the time scale that we can access with it, but there are techniques such as an efficient implementation of the reactive flow (RF) method, which is useful to treat the PBD model due to its character of first neighbor interaction. The development of multiscale DNA modeling has been carried out during the stay in the Applied Theoretical Chemistry group of the NTNU (Norwegian University of Science and Technology). Dr. Titus Sebastian van Erp's group is one of the world's foremost experts in DNA studies as well as in advanced modeling techniques. His techniques of acceleration and exploration of molecular reactions allowed us to study thermally activated DNA bubble singularities and DNA nanostructures in a complementary way thanks to the modeling, in synergy with the experimentation of neutrons under controlled pressure, temperature and atmosphere conditions. We have used techniques such as sampling of the transition interface (TIS) or an implementation of the reactive flux (RF) method [21] to develop our model.

The result of the simulations has been compared favorably with the experiments, verifying that the addition of the barrier improves the results of the original PBD model by several orders of magnitude. The implementation even reproduces the experimental data for short sequences although it gets worse the results as the length of the stem increases. That is to say, the proposed model with barrier, considerably improves the dynamics of DNA, especially for short sequences. We have played with the parameters of the model, the shape of the barrier and the height of the potential plateau, to see if the results improves. The conclusion has been that changing the depth of the Morse potential can reproduce all the experimental data checked, but this involves having unreal denaturation curves with melting temperatures above the boiling point of water. Thus, to obtain at the same time correct results of the dynamics and the thermodynamics of the DNA, improvements in the mesoscopic model are needed. It is necessary to add cooperative effect and consider the degrees of freedom that are missing. Finally, it has been seen that for dsDNA the model with the barrier also improves the description of the opening rates, but differs a little in the description of the distribution of bubbles and intermediary states.

For the dynamical studies using neutron scattering, high quality samples are necessary, therefore sample preparation and exhaustive characterization has been an essential part of this work. In Chapter 4 we have explained the methodology employed to obtain samples used in experiments as DNA films and artificial DNA short chains with controlled sequences.

In the last two years, our research group has been the first in the world to be able to synthesize DNA fibers in film structure and to characterize the structural correlations of their transition.

The synthesized DNA fibers have been used to carry out several experiments at the ILL facilities, for the study of the denaturation transition and the determination of how the persistence length changes through the melting transition correlating these changes with structural changes in the denaturing of DNA. For this purpose, in instruments D16 and D19, the evolution of the Bragg peaks of semi-crystalline and humidified DNA fibers was followed. A set of Li-DNA samples immersed in a water/ethanol mixture was studied varying the percentage from 20 to 80% ethanol and the results obtained have been compared with the statistical model PBD with an excellent agreement. Also the influence of spatial confinement on the structural correlation during the melting transition has been investigated in Na-DNA fibers immersed in solutions of poly (ethylene glycol) (PEG).

The precise synthesis has been achieved, in addition to developing a specific methodology for the design of short sequences of synthetic DNA, for its measurement by neutron scattering techniques. Several experiments have been carried out in particular at IN8 and D22 instruments (SANS- small angle neutron scattering) aimed at understanding the flexibility and rigidity of key sequences in DNA biological processes and their behavior with respect to temperature. Specifically, it has been proposed to obtain information about a sequence that is still a mystery in its biophysical bases, such as the sequence of Widom-601, involved in the positioning of histones [84].

Our results indicate the existence of kinked DNA molecules and suggest that, on its own, the local flexibility of DNA near the center of the "601" nucleosome positioning sequence is not enough to explain the positioning effect of the sequence. That is, in addition to the probable energetic effects in the interaction between histone and DNA, some type of effect or mechanical mechanism it could also contribute to positioning. Finally, the data obtained could contribute to the recent debate on the length scale dependence of DNA persistence length [132].

Finally, in Chapter 5, the melting transition in the synthesized DNA films was characterized by RAMAN microscopy through an exhaustive study and analysis of spectra of DNA fibers. In addition, a detailed physical characterization of these films has been made with advanced techniques such as optical microscopy (OM), scanning electron microscopy (SEM), and atomic force microscopy (AFM). This study has been completed with the analysis of DNA denaturation studied with RAMAN and the comparison of the melting transition of the dry fiber with DNA in solution and stabilized fiber at a higher relative humidity. Finally, the influence of molecular confinement on the transition of melting has been studied comparing the Raman spectra of the dry fiber with fibers submerged in different solutions of PEG and ethanol trying to see the distribution of these solvents in the fiber.

The first step was to determine melting profiles of DNA fibers using Raman spectroscopy. We have assigned bands present in the spectrum and we have analyzed the changes in the intensity of the bands and the shifts in the frequencies occurred due to an increment of temperature. With that we have obtained a denaturation curves that allow to visualize the melting transition. This profiles also allow to determine the width of the transition which gives information about the phase transition and its cooperativity. Our results are in good agreement with previous studies that report structural information, giving an average width of 5.97 ± 1.72 that suggests a smooth and continuous transition.

We also have done an analysis of the denaturation of DNA in solution and we have compared both transitions, in fiber and in solution. The comparison has been by quantification the shift of the frequencies occurred in the Raman bands between fiber and solution with the temperature. Again we have obtained good denaturation curves that allow us to calculate the transition width and compare both states, fiber and solution. Our results conclude that exist differences in the melting transition between both states where fiber present in all cases a smoother transition while solution exhibits sharper transition occurring in a narrow range of temperature indicating a cooperative strand separation, with a transition width of 2.98 ± 0.99. Finally, we have studied the solvent influence in DNA fibers. We have compared the Raman spectrum of the dry fiber with the ones obtained with the fiber submerged in different solvents such as water (relative humidity), poly(ethylene glycol) (PEG) or ethanol analyzing the molecular confinement on the fibers since PEG or ethanol can exerts osmotic pressure. Regarding fiber with higher relative humidity we have shown that the differences mainly arise because dry fiber is a mixture of A and B DNA form while fiber stabilized at 92% of RH is totally in B-form. This fact is reflected by the shifts of the frequencies of the main bands (especially those that have to do with the backbone and the formation of hydrogen bonds) of both fibers spectrums. As for fibers immersed in PEG and ethanol solutions, we have conclude that both solutions change the intermolecular distance between molecules (although PEG stabilizes DNA and ethanol destabilized the fiber), so both suffer changes in the frequencies related with structural changes (backbone and DNA form markers are affected) and with the elimination or formation of different hydrogen bonding. Fibers submerged in PEG or ethanol are swollen respect to dry fiber and the confinement of the fiber is reflected in the change of the intermolecular distance. By last we have tried to obtain 2D images selecting characteristics bands of both DNA and solvent (PEG or ethanol respectively) in order to study the distribution of the solution with respect to the fiber and our results suggest in both cases an homogeneous distribution.

6.2 Future Work

Until now, PBD model does not yet describe all features of DNA, but our work about dynamical properties such as opening rates of the molecule, have shown that it provides an improvement in the description of the. However, some essential alterations of the model must be made in order to obtain a correct description of the dynamics. One improvement that we are working on, is to replace the flat horizontal Morse plateau with a function having negative slope to describe entropic effects yet unconsidered such as rotational degrees of freedom and taking into account cooperative effects. To do that, it is necessary more accurate experimental data for fitting model parameters. We believe that this should be done on both thermodynamic and dynamical data. If all this suggesions are conducted this could lead to a much more reliable mesoscopic model for dsDNA and DNA hairpins.

Spatial constraints due to fiber structure should be further investigated, as it could play a role in DNA melting. More structural studies in oriented samples such as DNA fibers could provide additional information about structural changes during melting transition.

It should be necessary to do some improvements in the methodology followed to obtain synthetic DNA short chains in order to obtain a bigger amount of sample with higher quality launching a better purification protocol. In this way we can also avoid some troubles such as DNA aggregation finding in SANS experiments.

Regarding Raman experiments, we have on mind more comparative analysis such as the comparison of Li-DNA fibers (also used in neutron scattering experiments) with Na-DNA fibers, looking for differences in their spectrums that can suggest structural differences between this two types of fibers.

We have studied Na-DNA fiber submerged in a 60% of ethanol solution, but to do this experiment more realistic and in order to extract more feasible conclusions, we are measuring Raman spectrums of fibers immersed in ethanol solution with different concentration (80-50-35-20%).

All the comparative experiments with Na-DNA fiber by Raman spectroscopy should be extensible to Li-DNA fibers, comparing this fiber with Li-DNA in solution, and with fibers with high relative humidity, and submerged in PEG and ethanol solutions. We also want to study by Raman, dry fibers of Li/Na-DNA along the time varying the power of the laser in order to check some changes in the bands (intensity or shift) to correlate it with melting transition and structural changes. All comparative studies of fibers by Raman was done at room temperature so it would be interesting to repeat each analysis as a function of temperature, correlating changes in the melting due to this temperature increment.

A better physical characterization of the fiber samples is still needed. New SEM images of the fibers with different concentrations of PEG and ethanol solution would be really

interesting to elucidate physical changes that lead a differences in the melting transition. Also AFM images of each kind of samples are needed since only Na-DNA fiber has been studied.

6.3 Publications

The original content of this thesis appears in the following articles:

- Chapter 3:
  - Improving the mesoscopic modeling of DNA denaturation dynamics, M. Marty-Roda, O. Dahlen, T.S. van Erp, and S. Cuesta-López, Physical Biology (2018) (Under revision)
- Chapter 4:
  - Kinky DNA in solution: small angle scattering study of a nucleosome positioning sequence, A. González, T. Schindlerc, R. Boopathid, M. Marty-Roda, L. Romero-Santacreu, A. Wildes, L. Porcar, A. Martel, N. Theodorakopoulos, S. Cuesta-López, D. Angelov, T. Unruhc, and M. Peyrard, Nucleic Acids Research (2017) (Under revision)
  - The Melting Transition of Oriented DNA Fibers Submerged in Polyethylene Glycol Solutions Studied by Neutron Scattering and Calorimetry, A. Gonzáez, A. Wildes, M. Marty-Roda, S. Cuesta-López, E. Mossou, A. Studer, B. Deé, G. Moiroux, JL. Garden, N. Theodorakopoulos and M. Peyrard, J. Phys. Chem. B (Feb. 2018) (DOI: 10.1021/acs.jpcb.7b11608).
- Chapter 5:
  - Influence of three commercial graphene derivatives on the catalytic properties of a Lactobacillus plantarum alpha-L-rhamnosidase when used as immobilization matrixes, N. Antón-Millaá, J. García-Tojal, M. Marty-Roda, J. A. Tamayo-Ramos, S. Cuesta-López, ACS Applied Materials & Interfaces (2018) (Under revision).
  - Thermal oxidation of Ball-Milled ZnO doped powders for synthesis of nano-materials, E. Ferrone, S. Garroni, N. Antón- Millan, M. Marty-Roda 2 , S. Cuesta-López, M. Pea, A. Notargiacomo, A. Rinaldi, R. Araneo, IEEENANO
  - Dry DNA wet-spun films characterization and melting transition by RAMAN microscopy, M. Marty-Roda, A. Gonález, M. Peyrard, and S. Cuesta-López (In preparation)

– RAMAN Analysis of the solvent influence in the melting of DNA fibers, M. Marty-Roda, A. Gonález, M. Peyrard, and S. Cuesta-López (In preparation)

References

[1] A. Campa and A. Giansanti, Physical Review E **58**, 3585 (1998).

[2] A. Campa and A. Giansanti, Journal of biological physics **24**, 141 (1999).

[3] J. Jung and A. Van Orden, The Journal of Physical Chemistry B **109**, 3648 (2005).

[4] G. Bonnet, O. Krichevsky, and A. Libchaber, Proceedings of the National Academy of Sciences **95**, 8602 (1998).

[5] M. I. Wallace, L. Ying, S. Balasubramanian, and D. Klenerman, Proceedings of the National Academy of Sciences **98**, 5584 (2001).

[6] C. Hilbers, C. Haasnoot, S. De Bruin, J. Joordens, G. Van Der Marel, and J. Van Boom, Biochimie **67**, 685 (1985).

[7] A. Ansari, S. V. Kuznetsov, and Y. Shen, Proceedings of the National Academy of Sciences **98**, 7771 (2001).

[8] A. Ansari and S. V. Kuznetsov, The journal of physical chemistry B **109**, 12982 (2005).

[9] J. R. Grunwell, J. L. Glass, T. D. Lacoste, A. A. Deniz, D. S. Chemla, and P. G. Schultz, Journal of the American Chemical Society **123**, 4295 (2001).

[10] T. S. van Erp, S. Cuesta-Lopez, and M. Peyrard, Eur. Phys. J. E **20**, 421 (2006).

[11] T. Dauxois, M. Peyrard, and A. Bishop, Physical Review E **47**, 684 (1993).

[12] M. Peyrard and A. R. Bishop, Physical review letters **62**, 2755 (1989).

[13] J. D. Watson, F. H. Crick, *et al.*, Nature **171**, 737 (1953).

[14] R. A. Dimitrov and M. Zuker, Biophysical Journal **87**, 215 (2004).

[15] M. Sales-Pardo, R. Guimera, A. Moreira, J. Widom, and L. Amaral, Physical Review E **71**, 051902 (2005).

[16] S. A. García, *Modelos de superficies e intercaras: Transiciones de fase, desorden y aplicaciones*, Ph.D. thesis, Universidad Carlos III de Madrid (2005).

[17] J. V. Orero, *Dynamics and Thermal Behaviour of Films of Oriented DNA Fibres Investigated using Neutron Scattering and Calorimetry Techniques*, Ph.D. thesis, École Normale Supérieure de Lyon (2012).

[18] J. Errami, *Modelling DNA Hairpins*, Ph.D. thesis, École Normale Supérieure de Lyon (2007).

[19] M. Peyrard, S. Cuesta-Lopez, and D. Angelov, J. Phys.: Condens. Matter **21**, 034103 (2009).

[20] A. Wildes, N. Theodorakopoulos, J. Valle-Orero, S. Cuesta-Lopez, J.-L. Garden, and M. Peyrard, Phys. Rev. Lett. **106**, 048101 (2011).

[21] T. S. van Erp and M. Peyrard, EPL (Europhysics Letters) **98**, 48004 (2012).

[22] S. Cuesta-López, H. Menoni, D. Angelov, and M. Peyrard, Nucleic acids research **39**, 5276 (2011).

[23] N. R. Markham and M. Zuker, Nucleic acids research **33**, W577 (2005).

[24] M. Takinoue and A. Suyama, Small **2**, 1244 (2006).

[25] E. Ising, Zeitschrift für Physik A Hadrons and Nuclei **31**, 253 (1925).

[26] J. SantaLucia, Proceedings of the National Academy of Sciences **95**, 1460 (1998).

[27] D. Poland and H. A. Scheraga, The Journal of chemical physics **45**, 1464 (1966).

[28] D. Poland and H. A. Scheraga, The Journal of chemical physics **45**, 1456 (1966).

[29] D. Poland, Biopolymers **73**, 216 (2004).

[30] V. Ivanov, Y. Zeng, and G. Zocchi, Physical Review E **70**, 051907 (2004).

[31] C. Kittel, American Journal of Physics **37**, 917 (1969).

[32] N. Theodorakopoulos, T. Dauxois, and M. Peyrard, Physical Review Letters **85**, 6 (2000).

[33] W. Lenz, Z. Phys. **21**, 613 (1920).

[34] B. H. Zimm, The Journal of Chemical Physics **33**, 1349 (1960).

[35] S. Lieu, (2015).

[36] T. Dauxois, *Dynamique non linéaire et mécanique statistique d'un modèle d'ADN*, Ph.D. thesis, Universié de Bourgogne (1993).

[37] T. S. van Erp, S. Cuesta-Lopez, J.-G. Hagmann, and M. Peyrard, Physical review letters **95**, 218104 (2005).

[38] N. Theodorakopoulos and M. Peyrard, Physical review letters **108**, 078104 (2012).

[39] J. L. Leroy, M. Kochoyan, T. Huynhdinh, and M. Gueron, J. Mol. Biol. **200**, 223 (1988).

[40] J. Valle-Orero, A. R. Wildes, N. Theodorakopoulos, S. Cuesta-López, J.-L. Garden, S. Danilkin, and M. Peyrard, New J. Phys. **6**, 113017 (2014).

[41] E. Zazopoulos, E. Lalli, D. M. Stocco, and P. Sassone-Corsi, Nature **390**, 311 (1997).

[42] S. Froelich-Ammon, K. C. Gale, and N. Osheroff, Journal of Biological Chemistry **269**, 7719 (1994).

[43] T. Q. Trinh and R. Sinden, Genetics **134**, 409 (1993).

[44] D. J. Proctor, H. Ma, E. Kierzek, R. Kierzek, M. Gruebele, and P. C. Bevilacqua, Biochemistry **43**, 14004 (2004).

[45] J. Errami, M. Peyrard, and N. Theodorakopoulos, The European Physical Journal E: Soft Matter and Biological Physics **23**, 397 (2007).

[46] S. Cuesta-Lopez, M. Peyrard, and D. Graham, The European Physical Journal E: Soft Matter and Biological Physics **16**, 235 (2005).

[47] S. Tyagi and F. R. Kramer, Nature biotechnology **14**, 303 (1996).

[48] N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Physical review letters **85**, 2400 (2000).

[49] M. Takinoue and A. Suyama, Chem-Bio Informatics Journal **4**, 93 (2004).

[50] D. Liu and S. Balasubramanian, Angewandte Chemie International Edition **42**, 5734 (2003).

[51] Y. Chen, S.-H. Lee, and C. Mao, Angewandte Chemie International Edition **43**, 5335 (2004).

[52] J. Hanne, G. Zocchi, N. K. Voulgarakis, A. R. Bishop, and K. Ø. Rasmussen, Physical Review E **76**, 011909 (2007).

[53] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, Proceedings of the National Academy of Sciences **103**, 6190 (2006).

[54] T. S. van Erp, Physical review letters **98**, 268301 (2007).

[55] O. Dahlen and T. S. van Erp, The Journal of chemical physics **142**, 06B615_1 (2015).

[56] O. Kratky and G. Porod, "Rec trav chim pays-bas 1949, 68, 1106–1122," .

[57] T. S. Van Erp, "Dynamical rare event simulation techniques for equilibrium and nonequilibrium systems," in *Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials* (John Wiley & Sons, Inc., New York, USA, 2012) pp. 27–60.

[58] H. Eyring, The Journal of Chemical Physics **3**, 107 (1935).

[59] E. Wigner, Transactions of the Faraday Society **34**, 29 (1938).

[60] J. Keck, Discussions of the Faraday Society **33**, 173 (1962).

[61] J. Pu, J. Gao, and D. G. Truhlar, Chemical reviews **106**, 3140 (2006).

[62] J. Keck and G. Carrier, The Journal of Chemical Physics **43**, 2284 (1965).

[63] T. S. van Erp, The Journal of chemical physics **125**, 174106 (2006).

[64] R. W. Pastor, B. R. Brooks, and A. Szabo, Molecular Physics **65**, 1409 (1988).

[65] N. Theodorakopoulos, Physical Review E **82**, 021905 (2010).

[66] R. Inman and R. Baldwin, Journal of molecular biology **8**, 452 (1964).

[67] A. Montrichok, G. Gruner, and G. Zocchi, EPL (Europhysics Letters) **62**, 452 (2003).

[68] Y. Zeng, A. Montrichok, and G. Zocchi, Physical review letters **91**, 148101 (2003).

[69] Y. Zeng, A. Montrichok, and G. Zocchi, Journal of molecular biology **339**, 67 (2004).

[70] S. Ares and G. Kalosakas, Nano letters **7**, 307 (2007).

[71] T. Dauxois and M. Peyrard, Physical Review E **51**, 4027 (1995).

[72] Z. Rapti, A. Smerzi, K. Rasmussen, A. Bishop, C. Choi, and A. Usheva, Physical Review E **73**, 051902 (2006).

[73] A. Rupprecht, Biochemical and biophysical research communications **12**, 163 (1963).

[74] A. Rupprecht, Acta Chem Scand **20**, 494 (1966).

[75] A. Rupprecht, Biotechnology and bioengineering **12**, 93 (1970).

[76] A. Rupprecht and B. Forslind, Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis **204**, 304 (1970).

[77] A. Rupprecht, Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis **199**, 277 (1970).

[78] P. Schunk and L. Scriven, Journal of rheology **34**, 1085 (1990).

[79] A. Ziabicki, Colloid & Polymer Science **175**, 14 (1961).

[80] P. J. Hagerman, Annual review of biophysics and biophysical chemistry **17**, 265 (1988).

[81] J. P. Peters and L. J. Maher, Quarterly reviews of biophysics **43**, 23 (2010).

[82] M. Krisch, A. Mermet, H. Grimm, V. Forsyth, and A. Rupprecht, Physical Review E **73**, 061909 (2006).

[83] L. Van Eijck, F. Merzel, S. Rols, J. Ollivier, V. Forsyth, and M. Johnson, Physical review letters **107**, 088102 (2011).

[84] P. Lowary and J. Widom, Journal of molecular biology **276**, 19 (1998).

[85] D. Vasudevan, E. Y. Chua, and C. A. Davey, Journal of molecular biology **403**, 1 (2010).

[86] E. Y. Chua, D. Vasudevan, G. E. Davey, B. Wu, and C. A. Davey, Nucleic acids research **40**, 6338 (2012).

[87] J. D. Kahn, Biophysical journal **107**, 282 (2014).

[88] B. Hammouda and D. Worcester, Biophysical journal **91**, 2237 (2006).

[89] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. Mertens, P. V. Konarev, and D. I. Svergun, Journal of applied crystallography **45**, 342 (2012).

[90] M. Bastos, V. Castro, G. Mrevlishvili, and J. Teixeira, Biophysical journal **86**, 3822 (2004).

[91] D. Svergun, Journal of applied crystallography **25**, 495 (1992).

[92] O. Kratky and G. Porod, Recueil des Travaux Chimiques des Pays-Bas **68**, 1106 (1949).

[93] N. Theodorakopoulos, (2017).

[94] F. Crick and A. Klug, Nature **255**, 530 (1975).

[95] D. J. Gardiner, "Introduction to raman scattering," (1989).

[96] P. Graves and D. Gardiner, "Practical raman spectroscopy," (1989).

[97] C. V. Raman and K. S. Krishnan, Nature **121**, 501 (1928).

[98] V. S. Letokhov and V. P. Chebotayev, *Nonlinear laser spectroscopy*, Vol. 4 (Not Avail, 1977).

[99] N. Colthup, *Introduction to infrared and Raman spectroscopy* (Elsevier, 2012).

[100] Y. Shen, S. Yang, P. Zhou, Q. Sun, P. Wang, L. Wan, J. Li, L. Chen, X. Wang, S. Ding, *et al.*, Carbon **62**, 157 (2013).

[101] K. Novoselov and V. Fal, Nature **490**, 192 (2012).

[102] G. Binnig, Rec. Lett **56**, 9330 (1986).

[103] T. Albrecht, P. Grütter, D. Horne, and D. Rugar, Journal of Applied Physics **69**, 668 (1991).

[104] F. J. Giessibl, Materials Today **8**, 32 (2005).

[105] R. Reichelt, in *Science of microscopy* (Springer, 2007) pp. 133–272.

[106] C. Oatley, W. Nixon, and R. Pease, Advances in Electronics and Electron Physics **21**, 181 (1966).

[107] A. Wildes, N. Theodorakopoulos, J. Valle-Orero, S. Cuesta-López, J.-L. Garden, and M. Peyrard, Physical Review E **83**, 061923 (2011).

[108] A. Wildes, L. Khadeeva, W. Trewby, J. Valle-Orero, A. Studer, J.-L. Garden, and M. Peyrard, The Journal of Physical Chemistry B **119**, 4441 (2015).

[109] B. Prescott, W. Steinmetz, and G. Thomas, Biopolymers **23**, 235 (1984).

[110] J. Duguid, V. A. Bloomfield, J. Benevides, and G. Thomas, Biophysical journal **65**, 1916 (1993).

[111] J. G. Duguid, V. A. Bloomfield, J. M. Benevides, and G. Thomas, Biophysical journal **71**, 3350 (1996).

[112] L. Movileanu, J. M. Benevides, and G. J. Thomas, Biopolymers **63**, 181 (2002).

[113] J. M. Benevides, S. A. Overman, and G. J. Thomas, Journal of Raman Spectroscopy **36**, 279 (2005).

[114] S. C. Erfurth, P. J. Bond, and W. L. Peticolas, Biopolymers **14**, 1245 (1975).

[115] S. S. Chan, R. H. Austin, I. Mukerji, and T. G. Spiro, Biophysical journal **72**, 1512 (1997).

[116] S. C. Erfurth, E. J. Kiser, and W. L. Peticolas, Proceedings of the National Academy of Sciences **69**, 938 (1972).

[117] S. C. Erfurth and W. L. Peticolas, Biopolymers **14**, 247 (1975).

[118] I. Mukerji and A. P. Williams, Biochemistry **41**, 69 (2002).

[119] J. Valle-Orero, A. Wildes, J.-L. Garden, and M. Peyrard, The Journal of Physical Chemistry B **117**, 1849 (2013).

[120] S. P. Fodor, R. P. Rava, T. R. Hays, and T. G. Spiro, Journal of the American Chemical Society **107**, 1520 (1985).

[121] C. Demarco, S. Lindsay, M. Pokorny, J. Powell, and A. Rupprecht, Biopolymers **24**, 2035 (1985).

[122] Y. Tominaga, M. Shida, K. Kubota, H. Urabe, Y. Nishimura, and M. Tsuboi, The Journal of Chemical Physics **83**, 5972 (1985).

[123] J. M. Benevides, P. L. Stow, L. L. Ilag, N. L. Incardona, and G. J. Thomas Jr, Biochemistry **30**, 4855 (1991).

[124] C. H. Spink, Methods in cell biology **84**, 115 (2008).

[125] R. Langridge, H. Wilson, C. Hooper, M. H. Wilkins, and L. Hamilton, Journal of Molecular Biology **2**, 19IN9 (1960).

[126] J. Benevides and G. Thomas Jr, Nucleic acids research **11**, 5747 (1983).

[127] T. O'connor, S. Mansy, M. Bina, D. McMillin, M. Bruck, and R. Tobias, Biophysical chemistry **15**, 53 (1982).

[128] S. Yasar, R. Podgornik, J. Valle-Orero, M. R. Johnson, and V. A. Parsegian, Scientific reports **4** (2014).

[129] R. Podgornik, H. Strey, K. Gawrisch, D. Rau, A. Rupprecht, and V. Parsegian, Proceedings of the National Academy of Sciences **93**, 4261 (1996).

[130] C. H. Spink and J. B. Chaires, Biochemistry **38**, 496 (1999).

[131] P. Woolley and P. R. Wills, Biophysical chemistry **22**, 89 (1985).

[132] A. Noy and R. Golestanian, Physical review letters **109**, 228101 (2012).