Scuola di Dottorato in Fisica, Astrofisica e Fisica Applicata
Dipartimento di Fisica

Corso di Dottorato in Fisica, Astrofisica e Fisica Applicata
Ciclo XXVIII

# Exploring the physics of polymers and living polymers through DNA models

Settore Scientifico Disciplinare FIS/01

Supervisore: Prof. Tommaso Bellini

Coordinatore: Prof. Francesco Ragusa

Tesi di Dottorato di:
Marina Laura Rossi

Anno Accademico 2014-2015

**Commission of the final examination:**

External Referee:
Prof. Alberta Ferrarini

External Member:
Prof. Per Rudquist

Internal Member:
Prof. Stefano Zapperi

**Final examination:**

Date 14-06-2015

Università degli Studi di Milano, Dipartimento di Fisica, Milano, Italy

**Cover illustration:**

Liquid crystalline phases of a racemic mixture of a DNA 12mer.

**MIUR subjects:**

FIS/01

**PACS:**

87.14.-g

# Contents

**Introduction** **vi**
Thesis overview vii

**1 DNA, structure and interactions** **1**
1.1 DNA double helix 2
1.2 Interactions between nucleotides 4
1.3 Melting temperature 6

**2 Coil to globule transition of a single stranded sequence of DNA** **9**
2.1 Single-stranded DNA as a model for heteropolymers 11
2.2 Sample preparation 16
2.3 Polymer models 17
2.4 Dynamic Light Scattering of dilute polymer solutions 19
2.5 Diffusion of polymers in solution 21
2.6 DLS instrumentation 26
2.7 Fluorescence Correlation Spectroscopy 27
2.8 FCS apparatus 29
2.9 Results 30
2.10 Discussion 41
2.11 Conclusions 46

**3 Liquid crystals** **49**
3.1 Some liquid crystalline phases 49
3.2 Formation of a LC phases 52
3.3 LC phases of DNA 54
3.4 LC phases of short duplexes of DNA 55

**4 Liquid crystal ordering of B-DNA duplexes: extracting the interactions from the phase boundaries.** **61**
4.1 Sequence 63
4.2 Sample preparation 63
4.3 AllAT nematic phase 65
4.4 Order parameter 70
4.5 A coarse grained model for the DNA nematic phase 71

4.6 Determination of the shape and strength of the stacking interactions through combining atomistic simulations and the coarse grained model 81
4.7 Conclusions 100

**5 Matherials and methods 103**
5.1 Measurement of the melting temperature 103
5.2 LC optical anisotropy 103
5.3 Measurements of transmission spectra 107
5.4 Birefringence measurements 107

**List of Figures 109**

**List of Tables 115**

**Bibliography 117**

# Introduction

## Thesis overview

In the last twenty years DNA has become a topic of research in material science and technology. It was in fact realized, first by Nadrian Seeman [1] [2] and later by many other groups, that the code depending pairing of DNA polymers could be exploited to program their self assembly into complex multi-strand structures. This field of research, which is generally named DNA structure of nanotechnology, has been constantly growing. While at its origin the products of DNA assembly where simple geometrical shapes (nano cubes, nano tetrahedra, DNA origami [3] and so on), the latest achievements involve quite sophisticated programmable structures, including mobile nano-spiders, fueled by DNA hairpins and DNA Lego, in which the coded assembly of DNA is used to realize a toolbox of DNA Lego bricks of various size and shapes. The general philosophy underlying this effort is to try to master DNA properties to obtain complex nano-structures by spontaneous self-assembly. The aim of this field is, in principle, technological. Only in few cases applications were proposed: artificial ion channels crossing cell membranes and drug carriers that would open when an aptamer-protein interaction occurs. Despite the still unclear application of this technology, it is certainly true that the assembly of DNA enables producing nano-structures having a precision and complexity which is not achieved with any other molecular platform. Indeed, when any of these nanoscale structures is designed, it is clear which single base of which specific strand will end up in each position of the nano object.

In parallel to this enormous development of DNA nanotechnology, a distinct branch of material science research has emerged in which the properties of DNA are exploited, namely the use of DNA structures as a molecular model capable of setting the stage to explore relevant processes of condensed matter physics. Among these the use of DNA to explore the physics behind the formation of LC phases, and the use of DNA aggregates as a model of low valence molecules. In this branch of research, two specific properties of DNA are exploited. First, the mature capacity of synthesizing DNA oligomers at will enables to produce molecular solutions with a thermodynamically relevant quantities of identical structures. This is crucial when studying the phase behavior whose interpretation is made simpler when the particles are identical. Secondly, DNA offers the possibility of modulating inter-chain interactions. One peculiar and extremely important property of DNA is the strong temperature-dependence of hybridization interactions. Therefore by changing both length of the DNA chains involved in the structures and the temperature one can explore a huge range of intermolecular interactions with no analogues in other molecular solutions. Besides the interaction strengths, DNA also enables to build mutually interacting nano structures also controlling their lifetime [4].It is in fact well known that the lifetime of two DNA oligomers paired in a duplex enormously depend on the length of the complementary sequence and temperature. At room temperature the lifetime of an eight-base pair duplex is of the order of milliseconds, while the lifetime of a 12 base pair duplexes is of the order of hours. This is very well known in biology where the selectivity and stability of the pairing with PCR primers is daily exploited.

This thesis is part of this field of research in which open problems of condensed matter, and specifically of soft matter physics, are reproduced in a solution of DNA molecules. The class of problems here considered involves both polymer physics and liquid crystal formation. The latter is also intimately related to polymerization in that the formation of LC phases in short DNA oligomers is mediated by their aggregation into living polymers. Indeed, polymerization and LC formation, besides sharing a basic uniaxial symmetry, might be even more intimately connected as the formation of LC ordering might even be the cradle for the development of long polymeric chains obtained by the chemical stabilization of living polymers [5]. Specifically, during my PhD research activity I considered four distinct topics: aggregation and LC ordering of racemic mixtures of B-DNA duplexes, aggregation and LC formation of achiral DNA duplexes, living polymerization of strongly associating DNA duplexes with long complementary overhangs, globulization transition of a long single strand of DNA. Only the first and the last topics reached a level of maturity so to be enclosed in this final dissertation.

This thesis work is organized in five chapters.

> **Chapter 1: DNA, structure and interactions**. The principal characteristics of DNA are resumed as a tool to understand the following studied performed using DNA molecules. The chapter consists of a brief overview of the the structure of a strand and a double helix of DNA, and an explanation of the interactions that are responsible of the formation of the double helix.
>
> **Chapter 2: Liquid crystal ordering of B-DNA duplexes: extracting the interactions from the phase boundaries**. The second chapter is about the study of the coil to globule transition of polymers made through measuring the dimension of a long single strand of DNA in a dilute regime as a function the temperature. The coil to globule transition was studied since the early stages of polymer science and is still now a topic of research because the properties of the polymer are found to influence the transition process in a complex and still largely undetermined way. In particular, we chose a single strand of DNA as good model of heteropolymers because segments of the strands that are composed by consecutive complementary bases interact to form a segment of double helix. The number of bases involved and the distance along the chain of such segments give rise to different interaction strengths. We used two techniques, Dynamic Light Scattering and Fluorescence Correlation Spectroscopy, and with both we observed a transition of the single chain from a smallest to a largest dimension in a sharp temperature range. We also propose an interpretation of the thermodynamics of the possible interactions that could happen inside a chain.
>
> **Chapter 3: Liquid crystals** An brief introduction to the properties and kind of liquid crystalline (LC) phases is explained. Those phases are usually observed in a assemblies of elongated molecules and yield a partial ordering of the molecules. The conditions necessary to the formation of LC phases for elongated molecules were exhaustively studied by Onsager, that showed that the molecules should ex-

hibit enough shape anisotropy to form LC phases. Recently a LC ordering was found in solutions of molecules that does not possess enough shape anisotropy as Onsager theory predicted. Examples of such molecules are the chromonics or small duplexes of DNA, made of less than 20 base pairs. It was found that in this case the phenomenon that drives the formation of LC phases is the living polymerization: for example the ends of DNA duplexes are hydrophobic and as a consequence when the duplexes are diluted in aqueous solutions the ends of different molecules interact via stacking forces to form long linear aggregates that minimize the exposition of hydrophobic parts towards water.

**Chapter 4: Liquid crystal ordering of B-DNA duplexes: extracting the interactions from the phase boundaries.** This chapter is about the study of the stacking interaction between blunt ended DNA duplexes. As previously anticipated, the stacking interaction is responsible for the formation of the LC phases of short DNA duplexes, but up to now there is no experimental measurement of its strength between two base pairs at the terminals of two distinct duplexes. Some models were proposed to account for the onset of ordering in such systems; we referred to one simple model developed by De Michele et al. that allows an analytical calculation of the principal properties of the nematic phase. In this model the duplexes were represented by simple geometrical forms that interact through a sticky potential at the ends. Another approach was followed by Maffeo et al. to investigate the basis of the LC ordering. They performed atomistic simulations to extract the potential mean force between the ends of two duplexes as a function of their distance and relative orientation. We propose a method to give an estimate of the strength and the profile of the stacking interaction between phase pairs at the terminals of two different duplexes. Such method is based on two steps. The first consists on the characterization of a nematic phase observed in solutions of short duplexes. As a second step we compared the measured quantities to the prediction obtained by the combination of the coarse grained model with the atomistic simulations, to extract new information on the stacking interaction between independent helices.

**Chapter 5: Experimental methods**. In this chapter we explained the experimental techniques used to perform the measurements for the characterization of the nematic phase of short DNA duplexes. We mainly used microscopy based techniques.

CHAPTER 1

# DNA, structure and interactions

The deoxyribonucleic acid (DNA) is a polymer composed by monomers that are called nucleotides. Nucleotides consist of about 20 atoms, that form the deoxyribose (a sugar), the phosphate group and the nucleobase, like shown in Fig.1.1. Usually the carbon atoms of the sugar and phosphate moieties are indicated by numbers from 1 to 5. DNA is formed by four kind of nucleobases, called adenine, thymine, guanosine and cytosine. Adenine and guanosine are called purines and are bigger than the others, that are called pirimidines. Fig.1.2 shows the differences in composition and dimension between the four nucleobases.

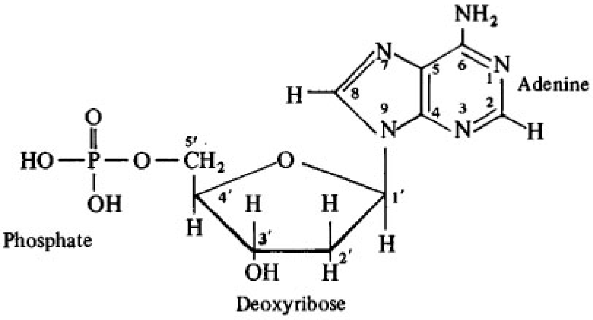


**Figure 1.1:** Sketch of a DNA nucleotide with the adenine as the nucleobase.[6]

The phosphodiesteric bond between the phosphate group and the sugar of the consecutive nucleotide (see Fig.1.3) creates a chain where the sugar is bonded to two phosphate groups at both $5'$ and $3'$ ends forming a long single-stranded DNA chain. Phosphate group and deoxyribose are found in all the DNA chains, that differentiates one from the other for the sequence of the nucleobases. As a consequence to characterize a DNA molecule is sufficient to list the sequence of nucleobases from the carbon atom in position $5'$ of the first sugar up to the nucleobases bond to the $3'$ atom of the last sugar. If two single stranded DNA chains are close they bind together forming a double helix of double-stranded DNA (dsDNA). They dispose in an antiparallel way: the $5'$ carbon atom of the first chain is near to the $3'$ of the second because this is the only way to bind two nucleobases. The double helix structure typical of dsDNA is shown in Fig.1.4

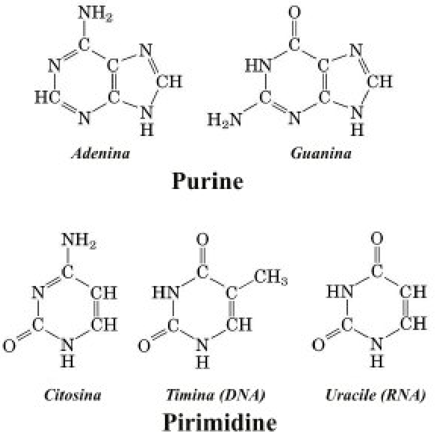


**Figure 1.2:** Different kind of nucleobases found in DNA and RNA molecules. Uracyle substitutes thymine in RNA molecules.

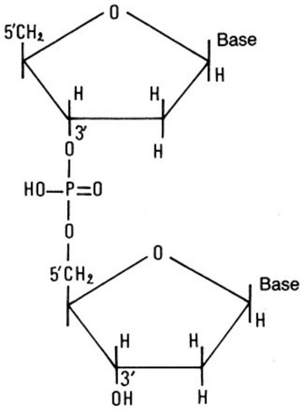


**Figure 1.3:** Phosphodiesteric bond between two nucleotides. Nucleobases, that should be found on the right side of sugar, are not shown.

## 1.1 DNA double helix

In physiological conditions two complementary sequences of DNA form a double helix, that has about $10-12$ couples of bases per turn and is characterized by a pitch of $3-4\,nm$ and a diameter of about $2\,nm$. The peculiarity of the structure can change due to the sequence of bases and the environmental conditions. In Fig.1.5 the three main structures of dsDNA and dsRNA are shown. Typically DNA is found in a *B-form* and seldom in an *A-form*, while RNA assumes a structure called *Z-form*. The helicity of the DNA structure gives a chirality to the molecule. Moreover, all the DNA helices observed in nature always have the same right-handed chirality.

DNA has another important property: it is a polyelectrolyte because every phosphate group has a negative charge of $-e$. When dsDNA assumes the helicoidal structure also

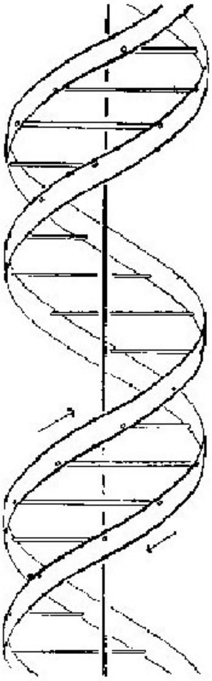


**Figure 1.4:** Sketch of a *double stranded*-DNA molecule

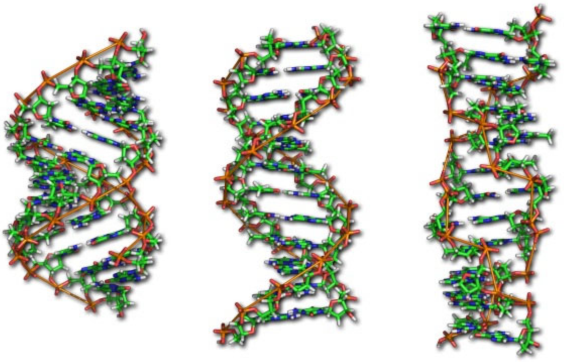


**Figure 1.5:** Different 3-D structures of double-stranded DNA and RNA. From left to right they are called *A-form*, *B-form* and *Z-form*.

the charges arrange in a helicoidal way, that is a chiral distribution. The linear density of charge of a double stranded DNA molecule is of about $-6\,e/nm$. If we dissolve a DNA double helix in a water solution with some salt concentration we observe that the positive ions from the salt dissociation (called counterions) come close to the double helix and provide a (partial) screening of the electric charge of the helix. The screening intensity depends on the ionic strength of the solution. The distance from the helix at which the electrostatic potential is decreased by a factor $1/e$ is the Debye screening length, that is proportional to the square root of the ion concentration in the solution.

## 1.2 Interactions between nucleotides

The nucleobases interact through two main forces: pairing and stacking forces. The pairing force is highly selective and allows thymine to bind with adenine and guanosine with cytosine. Pairing interaction is highly selective because the bond that form are hydrogen bond, like shown in Fig.1.6. Between adenine and cytosine we observe the formation of two hydrogen bonds, while between guanosine and cytosine the interaction is more intense because there are three bonds forming. The couples of bases that can bind through pairing interactions are called Watson-Crick couples from the names of the first that discovered the pairing interaction between the nucleobases.

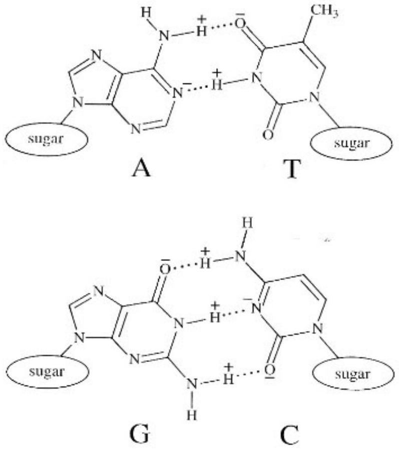


Figure 1.6: Possible pairing interactions between the nucleobases.

DNA nucleobases are highly hydrophobic, while the molecules of the side chain, sugars and phosphate groups, are hydrophilic. This is a factor that contributes to the formation of the double helix, that is a structure that allows the hydrophobic parts to be hidden from water but at the same time it leaves the hydrophilic parts close to water molecules. Nucleobases because of their hydrophobicity try to minimize the surface in contact with water and also to maximize the overlapping surface between them. This kind of interaction between the bases is called stacking interaction and it has been observed that its intensity depends on the overlap surface between the bases. It is generally known that tacking interactions are less selective than the pairing ones because no preferential stacking between couples of bases has been noticed up to now.

As we can see in Fig.1.7 stacking and pairing forces between the nucleobases have a different dependence on temperature: the stacking force strongly increases when the temperature is decreases, while the pairing force exhibits no variation with temperature [7]. From Fig.1.7 we can deduce that at all temperatures stacking forces are more intense

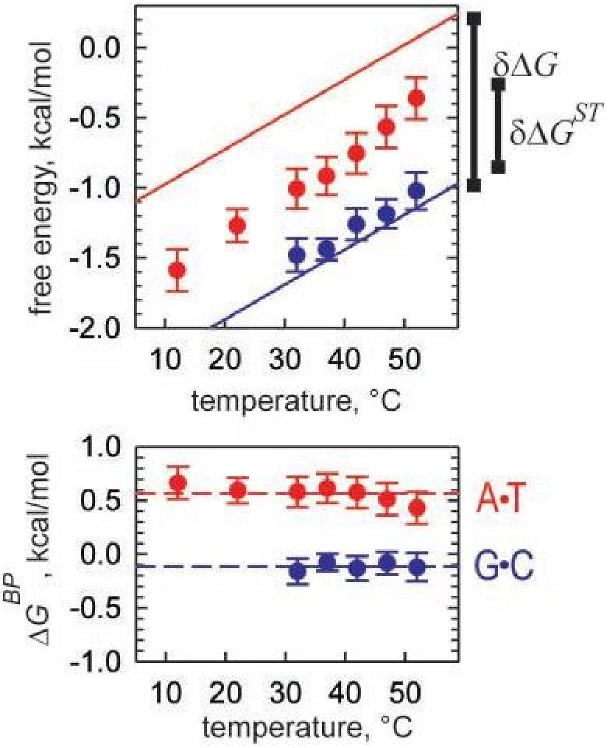


**Figure 1.7:** Stacking and pairing forces between various kind of nucleobases at different temperatures [7].

than pairing interactions. Their intensity of the stacking force depends on the nucleobases involved and also on the specific structure of the DNA molecules used to investigate the interactions. In chapter 4 stacking forces between DNA molecules are largely investigated. In the case of Fig.1.7 stacking interaction was investigated through the study of a nicked double helix of DNA, i.e. a helix without one phosphodiesteric bond and so without a stacking contribution. The pairing force was found to have a small intensity, and even not to be a favorable interaction, because the nucleobases could in principle form hydrogen bonds also with water. The double helix structure is then stabilized by stacking interactions.

The phosphate group and the deoxyribose are linked with a covalent bond. The double helix has strong bonds between molecules of the side chain and weak bonds between the bases and this fact allows the transcription and replication processes to happen. A DNA helix has to wind and unwind fastly to be functional. For example if we imagine to take a the human genome from chromosomes and to unwind it we would find a $2m$ long molecule. This molecule in the cell nucleus is winded around proteics groups called nucleosomes for an average of about $150$ bases around each nucleosome. Nucleosomes than are crowded to obtain the chromosomes that are about $10\,\mu m$ diameter. During transcription and replication processes the DNA molecule unwinds fastly from a very compact structure, and it also opens regions of the helix to copy them to another strand.

If we observe the two-dimensional structure of two ideally close single strands of

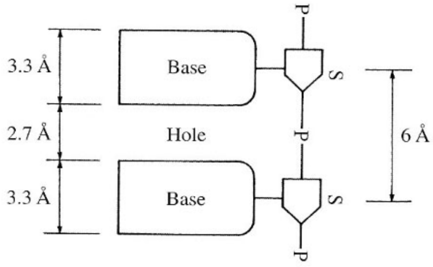


**Figure 1.8:** Consecutive nucleotides ideally put in the same bidimensional plane [8].

DNA, like in Fig.1.8, we notice that the spacing between the two bases is extremely larger than their thickness, so a space of about $2,7\ \AA$ would be available for water molecules to enter. One way to prevent such molecules to come close to the hydrophobic bases is to rotate the chain, like what happens for the double helix structure.

To characterized in a more precise way the double helix structure, we can notice that to minimize the surface exposed to water the bases rotate around the helical axes. Their position is defined through a set of three parameters, that are the twist angle, i.e. the angle of the couple of bases around the main axis of the helix and has a value of about $35°$ for the B-form, the roll angle that is the rotational angle around the axes of the couple of bases and the slide that describes the translation of one base from the previous one of the same strand. For B-form the average roll angle is in between the range $+20°$ and $-10°$, while the slide assumes values from the range between $3$ and $-2\ \AA$. In Fig.1.9 roll, twist and slide are shown.

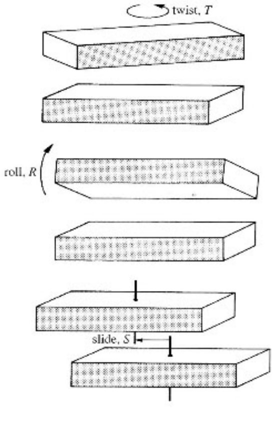


**Figure 1.9:** Twist, slide e roll of a double helix of DNA [8]

## 1.3 Melting temperature

The strength of the interactions of a DNA double helix can be measured through the melting temperature. For a double helix of DNA dissolved in a solution if the temperature is increased then the stacking forces become less and less intense until the helical structure becomes unstable and unfolds. The melting temperature $T_m$ of a solution of identical double strands of DNA at some ionic strength is defined as the temperature at which half of the helices are formed and half are unfolded. There are several methods that allows to measure the melting temperature, one of these is based on the hyperchromicity of DNA. DNA in fact exhibits a peak in the absorption spectrum for the electromagnetic radiation at $260\,nm$. The absorbance varies as the structure of the molecule is changed: the double helix absorbs about a $10\%$ radiation more than the two single strands, so a measure of the radiation absorbed as a function of the temperature of the solution provides an accurate estimate of the melting temperature. The melting temperature depends on the length of the DNA strands, the specific sequence of bases, the concentration of the DNA molecules in the solution and also on the ionic strength. In particular the ionic strength affects much the melting temperature and so it is a key ingredient for the stability of the double helix in solution.

## CHAPTER 2

# Coil to globule transition of a single stranded sequence of DNA

The conformational space of single polymers in solution depends both on the interactions between the polymer and the solvent and on the nature of the interactions between monomers such as their specificity and cooperativity. Upon modifying the interactions between the polymer and the solvent, the chain may undergo a transition between different conformational states. The coil to globule transition, which describes the transformation of the chain from a swollen to a compact state, was studied since the early stages of polymer science. Despite the numerous studies, it has remained a topic of current research. This is because the properties of the polymer are found to influence the transition process in a complex and still largely undetermined way.

When the solvent is good, i.e. it strongly binds with the polymer, or when the temperature is high the polymer adopts a swollen or coiled conformation, because the monomers interact preferentially with the solvent then with the other monomers and/or because the entropic advantage of the swollen state is dominant. If the quality of the solvent is lowered or the temperature is decreased the interactions between the monomers become dominant over that between monomers and solvent and the polymer assumes a more compact (globular) conformation. This notion is captured by the classical Flory theory that describes the coil to globule transition in a highly idealized polymer that changes its conformation while the properties of the solvent are changed. The main success of the Flory theory is of providing the scaling behaviour of the mass vs the dimension of the polymer in the coil and globular conformations with a few but effective assumptions. The Flory approach is based on an analytic expression of the free energy $F_N(R)$ of a polymer made by N monomers in a good solvent or high temperature limit as a function of the average end to end distance $R$. The end to end distance $\mathbf{R}_e$ of a single conformation of a chain is the distance between the positions of the first and the last monomer $\mathbf{R}_e = \mathbf{r}_N - \mathbf{r}_0 = \sum_{i=1}^{N} \mathbf{r}_i - \mathbf{r}_{i-1}$. When averaged over all the possible conformations it becomes $R = \sqrt{< \sum_{i,j=1}^{N} (\mathbf{r}_i - \mathbf{r}_{i-1}) \cdot (\mathbf{r}_j - \mathbf{r}_{j-1}) >}$. The Flory free energy is:

$$F_N(R) = e_0 \frac{R^2}{Nb^2} + e_1 v_{exc} \frac{N^2}{R^3} + F_0 \tag{2.1}$$

$e_0$, $e_1$ are two T-dependent constants and $v_{exc}$ is the excluded volume of each pair of monomers. The free energy consists of three terms: the first term is the entropy that accounts for the number of conformations available for a polymer with end to end distance R, the second term is the repulsive energy due to the excluded volume, while the third term takes into account the contributions of the attractive interaction. In the Flory approach $F_0$ is approximated through the concentration expansion where only the the second virial coefficient is retained. As such this approach is intrinsically of mean field kind.

In d dimensions the minimization of the free energy $F_N(R)$ respect to the end to end distance gives a dimension of the chain $R \sim bN^{\nu}$ where $\nu = \frac{3}{d+2}$. This is the scaling behaviour of the coil conformation. At low temperature or in bad solvents the polymer adopts a more compact and homogeneous conformation and thus $R \sim N^{1/d}$. The temperature at which the transition between the two regimes occurs is the theta temperature. In this condition the repulsive and attractive contributions of the free energy balance. As a consequence the polymer resembles an ideal chain. In these three regions it is possible to calculate the dimension of the chain expressed by the end to end distance, for a dimension of the space $2 \leq d \leq 3$:

$$R = bN^{\nu_\theta}\phi(z) \tag{2.2}$$

where $b$ is the monomer length, $N$ is the number of monomers in a chain, $\nu_\theta = \frac{2}{d+1}$, $z = |\alpha|N^{\phi}$ and $\phi = \frac{d-1}{d+1}$. $\alpha$ is a parameter that express the 2-body interactions between the monomers. It is positive when $T > T_\theta$, null if $T = T_\theta$ and negative otherwise. Although most of the original description of the $\theta$-point was given as a function of the number of monomers and solvent quality, the experimental observations necessarily involve constant number of monomers and variable temperature, the temperature dependence of R is also implicit in Flory theory, as it has been well described by Giacometti.

Accordingly for polymers made by a finite number of monomers the temperature behaviour of the size of the chain can be calculated as:

$$R/b = (T - T_\theta)^{\frac{1}{d+2}} N^{\nu_F} \tag{2.3}$$

when $N \gg |T - T_\theta|^{-1/\phi}$ and $T > T_\theta$.

$$R/b = (T_\theta - T)^{-\frac{1}{d}} N^{\nu_c} \tag{2.4}$$

when $N \gg |T - T_\theta|^{-1/\phi}$ and $T < T_\theta$.

$$R/b = N^{\nu_\theta} \tag{2.5}$$

when $N \ll |T - T_\theta|^{-1/\phi}$ or $T = T_\theta$

If $N \to \infty$ the conditions for equations 2.3 and 2.4 are always satisfied. Thus the two regimes are contiguous and the dimension of the chain changes continuously from the dimension typical of $T < T_\theta$ and the one of $T > T_\theta$. At $T = T_\theta$ the theta point is found as the singular inflection point of the function that describes the behaviour of the end to end distance vs temperature. When N is finite the singularity characterizing the $\theta$-point is lost: the two regions 2.3 and 2.4 are no longer contiguous. In between there is a crossover transition. Around $T_\theta$ the chain satisfies 2.5 which is temperature-independent, thus indicating that the dimension of the chain is constant and no inflection point is found [9].

Flory theory is a very idealized model so it is not straightforward to understand to which kind of polymers it can be applied to. It is expected and experimentally observed that long flexible homopolymers are good Flory polymers. However as the properties of the polymer departs from this condition Flory's behaviour is lost. For example if the range of the interactions between the monomers is short [10], so that one monomer does not interact with all the other monomers at the same way, then the coil to globule transition becomes discontinuous. An analogous failure of the Flory theory is observed if the compact conformation that the polymer adopts at low temperatures is a crystalline conformation with a high degree of order [11]. Indeed, as the structure of the globular state becomes more specific, ordered and structured, the globulization process becomes discontinuous. This appears to have an implication for heteropolymers in which the heterogeneities imply stronger or selective interactions between specific monomers and thus a structural globular state. A paradigmatic example is that of protein folding, typically discontinuous despite the limited chain size. In general, an understanding of the $\theta$-point in heterogeneous chain is currently missing.

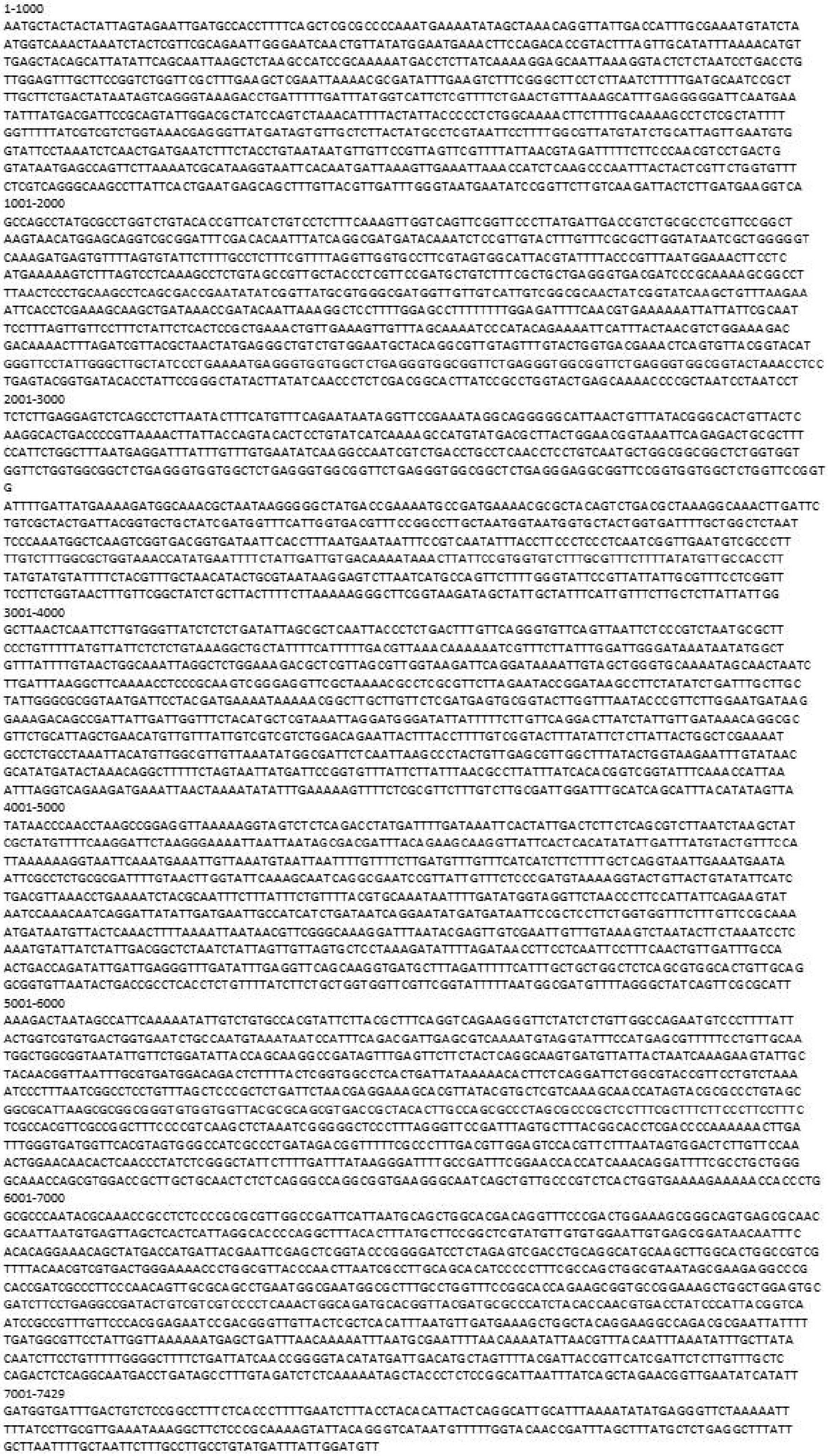
We investigated how the properties of a heteropolymer influence the behaviour of the coil to globule transition through the characterization of the conformational change of a single stranded DNA (ssDNA) in solution at different temperatures and ionic strengths.

ssDNA has been recently considered as a powerful tool to investigate the polymer dynamics and thermodynamics [12] [13] instead of the most commonly used double stranded (dsDNA) because of its greater flexibility. ssDNA has a smaller persistence length that dsDNA, about 0.6 nm compared to 50 nm, so even a short strand can be considered as a flexible chain. Moreover its structure is widely studied and well known, the interactions between monomers of the chain can be predicted with the simple Watson-Crick rules and we can easily obtain many copies of identical sequences. ssDNA can be considered as a heteropolymer because the interactions between the monomers are extremely specific, in fact two nucleobases can bind only if they satisfy the Watson-Crick pairing rules. In addition, ssDNA is characterized both long and short ranged interactions: the interaction between the nucleobases has a short range because it implies only the formation of hydrogen bonds, while the electrostatic repulsion that arises from the presence of a negative charge located on the phosphate group is a long range interaction.

## 2.1 Single-stranded DNA as a model for heteropolymers

As a model polymer we used a specific single stranded sequence of DNA composed by 7429 nucleotides. The sequence is the plasmidic genome of the M13mp18 virus that belongs to the family of the M13 bacteriophages. The genome of M13mp18 is modified from M13 with the insertion of *lacZ$\alpha$* gene and some cloning sites and is currently used in biochemical research. The complete nucleotide sequence of the genome of M13mp18 has been recently determined and is shown in Fig.2.1. (From now on M13mp18 will refer both to the virus and its genome.) The sequence was bought from Affymetrix and was shipped in a solution of 10mM Tris-HCl (pH 7.5), 1mM EDTA. Inside the virus the sequence is a circular plasmid, so we cut it to obtain a linear polymer with a procedure called linearization described in section 2.6.1. M13mp18 has already been used for DNA origami [3] because it is quite long and it does not have any strong secondary structure. If analyzed with the program DNA Mfold [3] it was found that the strongest secondary structure was a hairpin of just 20 bp (called Hairpin A) at positions 5515-5557 in the reference frame of Fig.2.1.

Because of the four different nitrogen bases involved in DNA sequence, a single strand of DNA is intrinsically a heteropolymer. As such it might have intra-chain interactions between portions of the chain that involve different strengths. The basic rule for such intra-chain interactions is given by the Watson-Crick pairing relations. According to these rules, adenine binds only with thymine and guanine with cytosine. However it is quite impossible that a bond involving just a couple of bases forms, because the energy gained through the pairing and stacking interactions is overwhelmed by the entropy loss due to the restriction of the conformational space of the two strands. To close the bond the two segment of sequences close to the complementary couples of bases must stay



```
1-1000
AATGCTACTACTATTAGTAGAATTGATGCCACCTTTTCAGCTCGCGCCCCAAATGAAAATATAGCTAAACAGGTTATTGACCATTTGCGAAATGTATCTA
ATGGTCAAACTAAATCTACTCGTTCGCAGAATTGGGAATCAACTGTTATATGGAATGAAACTTCCAGACACCGTACTTTAGTTGCATATTTAAAACATGT
TGAGCTACAGCATTATATTCAGCAATTAAGCTCTAAGCCATCCGCAAAAATGACCTCTTATCAAAAGGAGCAATTAAAGGTACTCTCTAATCCTGACCTG
TTGGAGTTTGCTTCCGGTCTGGTTCGCTTTGAAGCTCGAATTAAAACGCGATATTTGAAGTCTTTCGGGCTTCCTCTTAATCTTTTTGATGCAATCCGCT
TTGCTTCTGACTATAATAGTCAGGGTAAAGACCTGATTTTTGATTTATGGTCATTCTCGTTTTCTGAACTGTTTAAAGCATTTGAGGGGGATTCAATGAA
TATTTATGACGATTCCGCAGTATTGGACGCTATCCAGTCTAAACATTTTACTATTACCCCCTCTGGCAAAACTTCTTTTGCAAAAGCCTCTCGCTATTTT
GGTTTTTATCGTCGTCTGGTAAACGAGGGTTATGATAGTGTTGCTCTTACTATGCCTCGTAATTCCTTTTGGCGTTATGTATCTGCATTAGTTGAATGTG
GTATTCCTAAATCTCAACTGATGAATCTTTCTACCTGTAATAATGTTGTTCCGTTAGTTCGTTTTATTAACGTAGATTTTTCTTCCCAACGTCCTGACTG
GTATAATGAGCCAGTTCTTAAAATCGCATAAGGTAATTCACAATGATTAAAGTTGAAATTAAACCATCTCAAGCCCAATTTACTACTCGTTCTGGTGTTT
CTCGTCAGGGCAAGCCTTATTCACTGAATGAGCAGCTTTGTTACGTTGATTTGGGTAATGAATATCCGGTTCTTGTCAAGATTACTCTTGATGAAGGTCA
1001-2000
GCCAGCCTATGCGCCTGGTCTGTACACCGTTCATCTGTCCTCTTTCAAAGTTGGTCAGTTCGGTTCCCTTATGATTGACCGTCTGCGCCTCGTTCCGGCT
AAGTAACATGGAGCAGGTCGCGGATTTCGACACAATTTATCAGGCGATGATACAAATCTCCGTTGTACTTTGTTTCGCGCTTGGTATAATCGCTGGGGGT
CAAAGATGAGTGTTTTAGTGTATTCTTTTGCCTCTTTCGTTTTAGGTTGGTGCCTTCGTAGTGGCATTACGTATTTTACCCGTTTAATGGAAACTTCCTC
ATGAAAAAGTCTTTAGTCCTCAAAGCCTCTGTAGCCGTTGCTACCCTCGTTCCGATGCTGTCTTTCGCTGCTGAGGGTGACGATCCCGCAAAAGCGGCCT
TTAACTCCCTGCAAGCCTCAGCGACCGAATATATCGGTTATGCGTGGGCGATGGTTGTTGTCATTGTCGGCGCAACTATCGGTATCAAGCTGTTTAAGAA
ATTCACCTCGAAAGCAAGCTGATAAACCGATACAATTAAAGGCTCCTTTTGGAGCCTTTTTTTTGGAGATTTTCAACGTGAAAAAATTATTATTCGCAAT
TCCTTTAGTTGTTCCTTTCTATTCTCACTCCGCTGAAACTGTTGAAAGTTGTTTAGCAAAATCCCATACAGAAAATTCATTTACTAACGTCTGGAAAGAC
GACAAAACTTTAGATCGTTACGCTAACTATGAGGGCTGTCTGTGGAATGCTACAGGCGTTGTAGTTTGTACTGGTGACGAAACTCAGTGTTACGGTACAT
GGGTTCCTATTGGGCTTGCTATCCCTGAAAATGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTACTAAACCTCC
TGAGTACGGTGATACACCTATTCCGGGCTATACTTATATCAACCCTCTCGACGGCACTTATCCGCCTGGTACTGAGCAAAACCCCGCTAATCCTAATCCT
2001-3000
TCTCTTGAGGAGTCTCAGCCTCTTAATACTTTCATGTTTCAGAATAATAGGTTCCGAAATAGGCAGGGGGCATTAACTGTTTATACGGGCACTGTTACTC
AAGGCACTGACCCCGTTAAAACTTATTACCAGTACACTCCTGTATCATCAAAAGCCATGTATGACGCTTACTGGAACGGTAAATTCAGAGACTGCGCTTT
CCATTCTGGCTTTAATGAGGATTTATTTGTTTGTGAATATCAAGGCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGGCGGCTCTGGTGGT
GGTTCTGGTGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGAGGGAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGT
G
ATTTTGATTATGAAAAGATGGCAAACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTACAGTCTGACGCTAAAGGCAAACTTGATTC
TGTCGCTACTGATTACGGTGCTGCTATCGATGGTTTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGCTGGCTCTAAT
TCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCACCTTTAATGAATAATTTCCGTCAATATTTACCTTCCCTCCCTCAATCGGTTGAATGTCGCCCTT
TTGTCTTTGGCGCTGGTAAACCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTTCTTTTATATGTTGCCACCTT
TATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGGAGTCTTAATCATGCCAGTTCTTTTGGGTATTCCGTTATTATTGCGTTTCCTCGGTT
TCCTTCTGGTAACTTTGTTCGGCTATCTGCTTACTTTTCTTAAAAAGGGCTTCGGTAAGATAGCTATTGCTATTTCATTGTTTCTTGCTCTTATTATTGG
3001-4000
GCTTAACTCAATTCTTGTGGGTTATCTCTCTGATATTAGCGCTCAATTACCCTCTGACTTTGTTCAGGGTGTTCAGTTAATTCTCCCGTCTAATGCGCTT
CCCTGTTTTTATGTTATTCTCTCTGTAAAGGCTGCTATTTTCATTTTTGACGTTAAACAAAAAATCGTTTCTTATTTGGATTGGGATAAATAATATGGCT
GTTTATTTTGTAACTGGCAAATTAGGCTCTGGAAAGACGCTCGTTAGCGTTGGTAAGATTCAGGATAAAATTGTAGCTGGGTGCAAAATAGCAACTAATC
TTGATTTAAGGCTTCAAAACCTCCCGCAAGTCGGGAGGTTCGCTAAAACGCCTCGCGTTCTTAGAATACCGGATAAGCCTTCTATATCTGATTTGCTTGC
TATTGGGCGCGGTAATGATTCCTACGATGAAAATAAAAACGGCTTGCTTGTTCTCGATGAGTGCGGTACTTGGTTTAATACCCGTTCTTGGAATGATAAG
GAAAGACAGCCGATTATTGATTGGTTTCTACATGCTCGTAAATTAGGATGGGATATTATTTTTCTTGTTCAGGACTTATCTATTGTTGATAAACAGGCGC
GTTCTGCATTAGCTGAACATGTTGTTTATTGTCGTCGTCTGGACAGAATTACTTTACCTTTTGTCGGTACTTTATATTCTCTTATTACTGGCTCGAAAAT
GCCTCTGCCTAAATTACATGTTGGCGTTGTTAAATATGGCGATTCTCAATTAAGCCCTACTGTTGAGCGTTGGCTTTATACTGGTAAGAATTTGTATAAC
GCATATGATACTAAACAGGCTTTTTCTAGTAATTATGATTCCGGTGTTTATTCTTATTTAACGCCTTATTTATCACACGGTCGGTATTTCAAACCATTAA
ATTTAGGTCAGAAGATGAAATTAACTAAAATATATTTGAAAAAGTTTTCTCGCGTTCTTTGTCTTGCGATTGGATTTGCATCAGCATTTACATATAGTTA
4001-5000
TATAACCCAACCTAAGCCGGAGGTTAAAAAGGTAGTCTCTCAGACCTATGATTTTGATAAATTCACTATTGACTCTTCTCAGCGTCTTAATCTAAGCTAT
CGCTATGTTTTCAAGGATTCTAAGGGAAAATTAATTAATAGCGACGATTTACAGAAGCAAGGTTATTCACTCACATATATTGATTTATGTACTGTTTCCA
TTAAAAAAGGTAATTCAAATGAAATTGTTAAATGTAATTAATTTTGTTTTCTTGATGTTTGTTTCATCATCTTCTTTTGCTCAGGTAATTGAAATGAATA
ATTCGCCTCTGCGCGATTTTGTAACTTGGTATTCAAAGCAATCAGGCGAATCCGTTATTGTTTCTCCCGATGTAAAAGGTACTGTTACTGTATATTCATC
TGACGTTAAACCTGAAAATCTACGCAATTTCTTTATTTCTGTTTTACGTGCAAATAATTTTGATATGGTAGGTTCTAACCCTTCCATTATTCAGAAGTAT
AATCCAAACAATCAGGATTATATTGATGAATTGCCATCATCTGATAATCAGGAATATGATGATAATTCCGCTCCTTCTGGTGGTTTCTTTGTTCCGCAAA
ATGATAATGTTACTCAAACTTTTAAAATTAATAACGTTCGGGCAAAGGATTTAATACGAGTTGTCGAATTGTTTGTAAAGTCTAATACTTCTAAATCCTC
AAATGTATTATCTATTGACGGCTCTAATCTATTAGTTGTTAGTGCTCCTAAAGATATTTTAGATAACCTTCCTCAATTCCTTTCAACTGTTGATTTGCCA
ACTGACCAGATATTGATTGAGGGTTTGATATTTGAGGTTCAGCAAGGTGATGCTTTAGATTTTTCATTTGCTGCTGGCTCTCAGCGTGGCACTGTTGCAG
GCGGTGTTAATACTGACCGCCTCACCTCTGTTTTATCTTCTGCTGGTGGTTCGTTCGGTATTTTTAATGGCGATGTTTTAGGGCTATCAGTTCGCGCATT
5001-6000
AAAGACTAATAGCCATTCAAAAATATTGTCTGTGCCACGTATTCTTACGCTTTCAGGTCAGAAGGGTTCTATCTCTGTTGGCCAGAATGTCCCTTTTATT
ACTGGTCGTGTGACTGGTGAATCTGCCAATGTAAATAATCCATTTCAGACGATTGAGCGTCAAAATGTAGGTATTTCCATGAGCGTTTTTCCTGTTGCAA
TGGCTGGCGGTAATATTGTTCTGGATATTACCAGCAAGGCCGATAGTTTGAGTTCTTCTACTCAGGCAAGTGATGTTATTACTAATCAAAGAAGTATTGC
TACAACGGTTAATTTGCGTGATGGACAGACTCTTTTACTCGGTGGCCTCACTGATTATAAAAACACTTCTCAGGATTCTGGCGTACCGTTCCTGTCTAAA
ATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCTAACGAGGAAAGCACGTTATACGTGCTCGTCAAAGCAACCATAGTACGCGCCCTGTAGC
GGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTC
TCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGA
TTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA
ACTGGAACAACACTCAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGAACCACCATCAAACAGGATTTTCGCCTGCTGGG
GCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTG
6001-7000
GCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAAC
GCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTC
ACACAGGAAACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCG
TTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCG
CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGC
GATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGACCTATCCCATTACGGTCA
ATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTT
TGATGGCGTTCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAATGCGAATTTTAACAAAATATTAACGTTTACAATTTAAATATTTGCTTATA
CAATCTTCCTGTTTTTGGGGCTTTTCTGATTATCAACCGGGGTACATATGATTGACATGCTAGTTTTACGATTACCGTTCATCGATTCTCTTGTTTGCTC
CAGACTCTCAGGCAATGACCTGATAGCCTTTGTAGATCTCTCAAAAATAGCTACCCTCTCCGGCATTAATTTATCAGCTAGAACGGTTGAATATCATATT
7001-7429
GATGGTGATTTGACTGTCTCCGGCCTTTCTCACCCTTTTGAATCTTTACCTACACATTACTCAGGCATTGCATTTAAAATATATGAGGGTTCTAAAAATT
TTTATCCTTGCGTTGAAATAAAGGCTTCTCCCGCAAAAGTATTACAGGGTCATAATGTTTTTGGTACAACCGATTTAGCTTTATGCTCTGAGGCTTTATT
GCTTAATTTTGCTAATTCTTTGCCTTGCCTGTATGATTTATTGGATGTT
```

**Figure 2.1:** Sequence of basis of M13mp18 ssDNA from 5′ to 3′. The starting point is arbitrary.

close and then feel mutual steric and electrostatic forces due to the presence of the other sequence. In addition the closure of a bond between two regions of the same sequence introduces a loop inside the chain that decreases the entropy of the chain upon closing the bond. If there are two consecutive bases that are complementary to another consecutive couple of bases at some distance on the chain then the energetic gain increases because the number of bases involved in the bond increases, while the entropy loss remains almost constant because the either the restriction of the conformational space of the two sequences change slightly from the previous scenario and also the entropy loss caused by the formation of the loop stays unchanged. If there is a region of the chain where there are enough consecutive bases that are complementary to an equal number of consecutive bases located in a different region, then a segment of double helix may form. By borrowing a definition developed within the context of protein folding, two regions of the chain that bind are called contact regions. Every possible contact (bond) has an associated free energy that depends on the number of nucleobases involved and the distance between the two complementary contact regions along the sequence (i.e. the number of monomers between the two regions).

To characterize the ssDNA sequence used, we thus studied the statistics of the number of possible contacts as a function of the number of bases involved. To find every possible contact inside the ssDNA sequence we built a simple tool using the software Mathematica. In order to do that we considered the interactions of a sequence with its anti-parallel replica, at all possible relative positions by sliding one along the other. We looked for the possible complementary regions between the two. As the starting configuration the 5′ end of the sequence (A) was coupled with the 3′ end of the replica (B) to see if they were complementary. Then the replica was slid of one base in order to couple the 3′ end of B with the second base of A and the second base of B with the 5′ end of A, The procedure was repeated until only the 5′ end of sequence B was overlapping with 3′ end of sequence A. At each relative position we checked for complementary sequences. As some complementarity was found, we recorded its positions and the number of bases involved. On the basis of what discussed above we considered only complementary sequences of more than five base pairs. Then we listed all the contacts found and we checked if one or more bases were involved in more than one. In that case we considered just the paring that involves the larger number of bases. In Fig.2.2 we plotted the number of contacts found as a function of the number of bases involved in each contact. The maximum number of bases per contact that we found is 12, while the number of contact points increases when the number of bases involved decreases.

To understand whether the M13mp18 sequence has some transcription selected patterns that enhance or decrease the level of self-interactions, we compared this sequence with an analogous study was done on 4 random sequences, designed with the same percentage quantity of each kind of bases and of the same length of M13mp18. M13mp18 has more possible contacts that involve a number of bases greater than 9 than the random sequence, but over all, the behaviour of the number of bonds vs the number of bases involved in M13mp18 is similar to the one obtained with random sequences.

We studied also the contact map of M13mp18 compared to one of the random sequences. In Fig.2.3 we indicated the position of the monomers of the strand along both axes and we plotted each possible intra-chain contact at the coordinates individuated by the positions of the two contact regions. Different colors code for different number of bases involved in the contact. As in Fig.2.2 M13mp18 has some more possible contacts that involve a large number of bases (more than nine) then any of the random sequence studied. The contacts appear to be uniformly distributed along the chain for both sequences. Indeed, we could not identify patterns in the contact map that indicate that

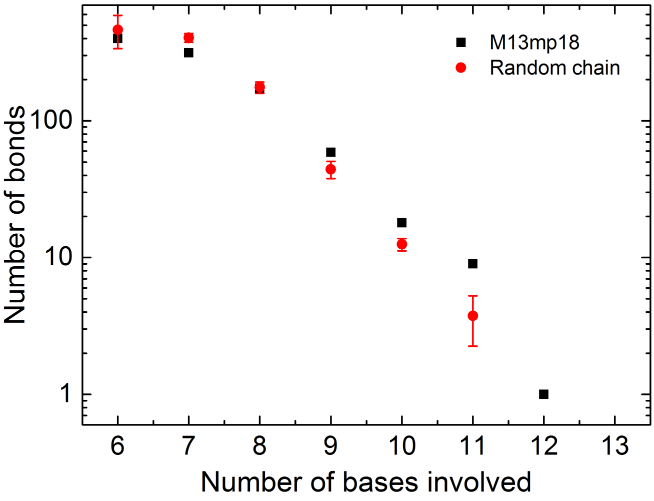


**Figure 2.2:** Number of possible contacts as a function of the number of bases involved in each contact for M13mp18 and for an average between 4 different random sequences with the same statistical distribution of bases as M13mp18. The error bars are the standard deviation of the number of contacts over the 4 random sequences considered. We neglected the contacts that involve less that 5 base pairs.

M13mp18 is significantly different from a random sequence of DNA.

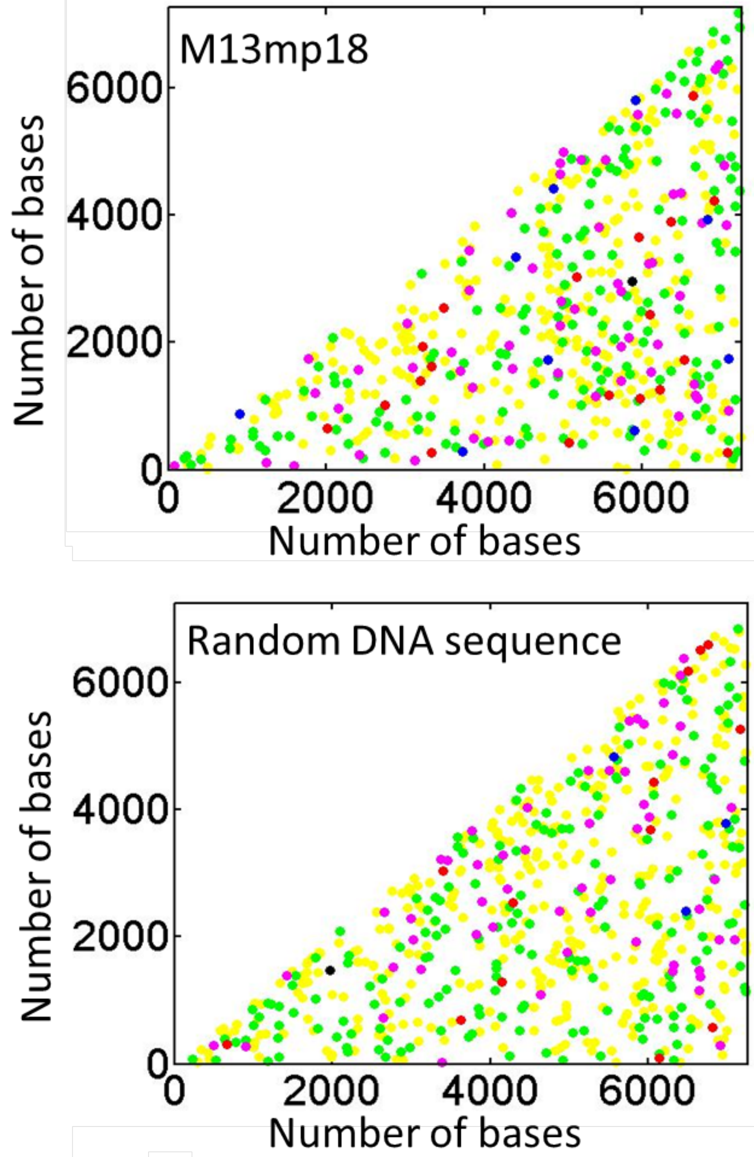


**Figure 2.3:** Contact map of M13mp18 (picture above) and for a random ssDNA sequence (picture below) with the same statistical distribution of bases as M13mp18.

## 2.2 Sample preparation

The conformation of M13mp18 extracted from the virus is circular. To model it into a linearized polymer it can be cut with an enzyme to become a linear polymer. The sequence has a recognition site for the enzyme BsRb1 [3] between the bases 5571 and 5576. The enzyme works at 37°C leaving a blunt end between 5573 and 5574 if it finds a double stranded region around those bases. To have a double stranded region we annealed the circular ssDNA with a short specific ssDNA sequence (primer). The primer has to be complementary to a segment of sequence of M13mp18 so that they could hybridize and form a segment of double helix. The binding energy depends on the length of the primer and also on temperature, so we chose a long enough primer to be sure to find a double stranded DNA at the recognition site at the working temperature of the enzyme. We also chose to include in the primer a sequence of bases complementary to a piece of Hairpin A, in order to prevent this structure to form.

We bought two different primers from Biomers because we used two techniques to measure the coil to globule transition, Dynamic Light Scattering (DLS) and Fluorescence Correlation Spectroscopy (FCS). The first was used to cut the samples for DLS, and it was complementary to the sequence 5560-5590 of M13mp18. After the cut it left two double stranded regions of 14 and 16 base pairs that had melting temperatures of respectively 40 and 60°C in our experimental condition. We got rid of the pieces of double helix by heating the sample at 90°C before the measurement so that the two halves of the primer would part from M13mp18. This procedure does not affect the scattering measurement because the scattered intensity of a sample is proportional to the square of the mass. If the sample consists of a mixture of short and long DNA sequences that have a mass about 500 times bigger then the scattered intensity of the shorter is negligible.

A different primer was needed for FCS measurements because we had to attach a fluorophore (Atto532, from Biomers) to the ssDNA complementary to the region 5549-5582 of M13mp18. The length of the primer was chosen as a compromise between the highest possible melting temperature from M13mp18 and the smaller possible length. In fact when the primer binds to the ssDNA, a piece of double helix is formed. This fragment of helix should be kept small because its dimension might affect the hydrodynamic radius of the linear DNA. We must also consider that if there is a fluorophore attached to the primer, the melting temperature of the primer from the ssDNA should be high to allow measurements of the hydrodynamic radius in the widest possible range of temperatures. Using the chosen primer the piece of double strand containing the fluorophore that forms after the annealing was of 20 base pairs. Its melting temperature is about 60°C.

For both DLS and FCS sample preparations, we annealed the primer to the circular DNA at 95°C for 2 min, then we decreased the temperature to 37°C at 0,2°C/min. The ratio between the number of primers and the ssDNA sequences in the cutting solution was 3:1. At 37°C we added the enzyme and its buffer (bought from EuroClone) and we waited for $2,5$ hours, as specified from the standard cutting procedure with BsRb1. Then we brought the solution at room temperature and we made a purification of the sample in two ways. For the sample for FCS we used the Agencourt AMPure Xp kit of magnetic beads (usually employed in PCR purification) while for the sample for DLS we filtrated the solution with Eurogold plasmid miniprep kit (bought from EuroClone). The first procedure assures a good removal of the primers but leaves some beads inside the sample, so it can be used for samples that have to be observed with fluorescence because the beads do not have fluorescent properties. The second procedure yields a worse removal of the primers but does not contaminate the sample, so it was used for scattering measurements because the signal of the primers is negligible compared to the signal of the ssDNA.

## 2.3 Polymer models

In this section some notions about the most simple models used to predict the polymer dimension are shown, because they constitutes the basis of our interpretation of the experimental data. A polymer is a chain made by a large number of the same repeating units, called monomers, or of different but resembling units. Several coarse-grained geometrical models are being used to predict how various physical quantities depend on the properties of the polymer as the chain length and the polymer concentration.

### 2.3.1 Ideal chain: the Random Walk

The simplest model of a polymer chain is the random walk. We place every monomer one after the other on a square lattice with a spacing between the sites that is equal to the distance between the monomers. If visiting the same site is allowed, the trajectory of the random walker is a model for an ideal chain. If not allowed, the trajectory resembles a real chain. For ideal chains we can assess the dimension and the size of the polymer through a couple of quantities. The end to end distance $R_F$ is the root mean square of the distance between the positions of the first and the last monomers of the chain.

$$R_F = \sqrt{< (\mathbf{r}_N - \mathbf{r}_0)^2) >} \tag{2.6}$$

where N is the total number of monomers. If we define the distance between two monomers as b we obtain

$$R_F = \sqrt{b^2 N} \tag{2.7}$$

The gyration radius is defined as the mean square of the distance between the monomers and the center of mass of the polymer.

$$R_G = \sqrt{\frac{1}{N+1} \sum_{i=0}^{N} < (\mathbf{r}_i - \mathbf{r}_G)^2 >} = \sqrt{\frac{1}{6} b^2 N} \tag{2.8}$$

Both quantities give a rough estimate of the dimension of the chain, though the $R_F$ can be defined only for linear chains, while $R_G$ can be calculated also for branched architectures.

We can introduce another parameter, that is the bond angle between adiacent monomers and we obtain the freely rotating chain. If we fix the angle at a value $\pi - \theta$ (see Fig.2.4) the end to end distance becomes

$$R_F = \sqrt{b^2 N \frac{1 + Cos(\theta)}{1 - Cos(\theta)}} \tag{2.9}$$

For smaller angles between the monomers the chain becomes stiffer and the end to end length increases. The gyration radius of the freely rotating chain is $R_G = \frac{1}{\sqrt{6}} R_F$

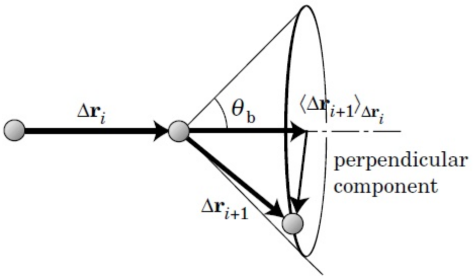


**Figure 2.4:** Two adiacent monomers of a chain with a fixed angle between them. [14]

### 2.3.2 Ideal chain: the Gaussian chain

The Gaussian chain is an idealized statistical description of a polymer chain where given a monomer in a position $\mathbf{r}_1$, the probability of finding another monomer at a position $\mathbf{r}_2$ and with a distance of n monomers along the chain is

$$G(\mathbf{r}_1, \mathbf{r}_2; n) = (2\pi nb^2/3)^{-3/2} exp\left(-\frac{3(\mathbf{r}_1 - \mathbf{r}_2)^2}{2nb^2}\right) \tag{2.10}$$

The end to end distance $R_F$ and the gyration radius $R_G$ are the same as for the ideal random walk. A Gaussian chain should be viewed as a hypothetical chain with the end-to-end distance where, when coarse-grained into N segments, each segment follows a Gaussian distribution with the mean square end-to-end length of ${R_F}^2/N$. The Gaussian chain has an unrealistic property: in 2.10, $\mathbf{r}_1$ and $\mathbf{r}_2$ can be separated by more than $nb$, although its probability is low when n is large. Despite this shortcoming, the Gaussian chain is the most preferred model in calculating various physical quantities in theories. It often happens that we can obtain an explicit analytic expression for the quantity in question only in the Gaussian chain model. It is the only model that gives an exact yet simple expression for the density of the chain ends, for instance.

### 2.3.3 Real chain: the excluded volume

A missing ingredient in the models seen so far is the excluded volume between monomers. This volume does not coincide with the volume occupied by the monomers, but is bigger. For simplicity, if we think at two identical rigid spheres in a solution we can calculate the excluded volume for a sphere due to the presence of the other one. It is eight times the volume of the sphere. The excluded volume introduces a variation in the entropy of the system because it reduces the number of configurations of the chain. Flory calculated the free energy of a chain where the entropy took into account the possible conformations of the chain and the energy depended on the interactions between different monomers. The dimension of the chain that minimizes the free energy is the gyration radius

$$R_G = bN^{\nu} \tag{2.11}$$

where $\nu = 3/5$. The $R_G$ does not depend on temperature so the chain is called athermal.

### 2.3.4 Semirigid chain: the wormlike model

There are many examples of semirigid chains amongst polymers, mainly due to the geometry of the bonds between monomers,the presence of side groups and electrostatic repulsion. A way to model semi-rigid chains is the Wormlike chain. It resembles the freely rotating chain but with an angle between monomers close to $\pi$ and with a short distance between monomers. If $\theta << 1$ and $b \to 0$ we can define a persistence length $L_p$ as the length along the chain after which the correlation between the orientation of the monomers is lost. If $\mathbf{u}(s)$ and $\mathbf{u}(s')$ are the vectors that indicate the direction of two monomers at a distance of respectively s and s′ monomers from the beginning of the chain, $L_p$ is defined as

$$< \mathbf{u}(s) \cdot \mathbf{u}(s') >= exp(-|s-s'|/L_p) \tag{2.12}$$

$$L_p = \frac{2b}{\theta} \tag{2.13}$$

The end to end distance and the gyration radius depend on the contour length of the chain $L_c = bN$ where N is the number of monomers and on the persistence length:

$$R_F = \sqrt{2L_p[L_c + L_p(exp(-L_c/L_p) - 1)]} \tag{2.14}$$

$$R_G = \sqrt{\frac{1}{3}L_pL_c - L_p^2 + 2\frac{L_p^3}{L_c}\left(1 - \frac{L_P}{L_c}[1 - exp(-L_c/L_p)]\right)} \tag{2.15}$$

## 2.4 Dynamic Light Scattering of dilute polymer solutions

Motions of polymer molecules in solution can be conveniently studied by using Dynamic Light Scattering (DLS). Measurements at a single scattering angle give information on the dimension of the polymer molecule in the solution with reasonable accuracy.

A typical scattering apparatus sends a laser source onto a sample, which for example consists of particles macromolecules or molecular aggregates dissolved in a solution, collects the light scattered in a given direction and detects the light diffused from the particles, when the light is collected for a single or a few coherence areas (corresponding to speckles). The diffused light $I(t)$ is not constant over time because of the Brownian motions of the particles (while the solvent molecules if homogeneous can usually be neglected). $I(t)$ can be written as a sum of an average and a fluctuating components:

$$I(t) =< I > +\Delta I(t) \tag{2.16}$$

The average component obviously does not change in time. To gain information from the kinetics of $I(t)$ the autocorrelation of the intensity over time is defined as in 2.17.

$$\langle I(t')I(t'+t)\rangle / \langle I\rangle^2 = \lim_{T_A\to\infty} \frac{1}{T_A\langle I\rangle^2}\int_0^{T_A} I(t')I(t'+t)dt = 1 + f_c g_2(t) \tag{2.17}$$

where $< I >$ is the average value of the intensity and $f_c = \langle \Delta I^2\rangle / \langle I\rangle^2$ is the coherence factor, whose value is about $0,5$ in our experimental conditions. We assume that the long-time average is equal to the ensemble average, i.e. the average with respect to the configuration of the system or, simply put, the average over all possible positions and shapes of the molecules in the solution. This assumption corresponds to assume ergodicity.

From 2.17 the normalized intensity autocorrelation function $g_2(t)$ is defined as $g_2(t) = \langle \Delta I(t')\Delta I(t'+t)\rangle / \langle \Delta I^2 \rangle$. It is possible to demonstrate that $g_2(t)$ is related to $g_1(t)$, a more basic property of the scattering process, through $g_2(t) = |g_1(t)|^2$, where

$$g_1(t) = \frac{< \mathbf{E}_s^*(t') \cdot \mathbf{E}_s(t'+t) >}{< \mathbf{E}_s^*(t') \cdot \mathbf{E}_s(t') >} \tag{2.18}$$

To understand the behaviour of $g_2(t)$ and $g_1(t)$ over time it is useful to define the dynamic structure factor of a suspension of particles:

$$S(\mathbf{k},t) = \left\langle \frac{1}{n_P} \sum_{m,n=1}^{n_P} exp[i\mathbf{k} \cdot (\mathbf{r}_m(0) - \mathbf{r}_n(t))] \right\rangle \tag{2.19}$$

where $n_p$ is the total number of molecules in the solution and $\mathbf{k}$ is the scattering vector $\mathbf{k} = \mathbf{k}_s - \mathbf{k}_i$, with $|\mathbf{k}| = \frac{4\pi n_{sol}}{\lambda} sin\left(\frac{\theta}{2}\right)$ where $n_{sol}$ is the refractive index of the solvent and $\theta$ the scattering angle. The dynamic structure factor depends on both the shape of the particles that compose the sample and also on the interactions between them. $S(\mathbf{k},t)$ can be considered as composed by two terms, one that is the single particle dynamic structure factor ($S_1(\mathbf{k},t)$) and the other that comes from the interactions between the particles.

$$S(\mathbf{k},t) = S_1(\mathbf{k},t) \times < (n_P - 1) exp[i\mathbf{k} \cdot (\mathbf{r}_1(0) - \mathbf{r}_2(t))] > \tag{2.20}$$

where

$$S_1(\mathbf{k},t) =< exp[i\mathbf{k} \cdot (\mathbf{r}_1(0) - \mathbf{r}_1(t))] > \tag{2.21}$$

It can be easily demonstrated that the field correlation function is connected to $S(\mathbf{k},t)$ by the 2.22

$$|g_1(t)| = \frac{S(\mathbf{k},t)}{S(\mathbf{k},0)} \tag{2.22}$$

$S_1(\mathbf{k},t)$ can be rewritten as a function of the transition probability $P(\mathbf{r},\mathbf{r}';t)$ that is the probability for a particle to move from $\mathbf{r}$ to $\mathbf{r}'$ in the time interval $t$.

$$S_1(\mathbf{k},t) = \int_V d\mathbf{r} exp[i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')] P(\mathbf{r},\mathbf{r}';t) \tag{2.23}$$

The sum is made over all the scattering volume and thus over all the particles.

### 2.4.1 Diffusion of particles in solution

Particles suspended in a liquid change their positions by diffusion. If a particle is much larger compared with the solvent molecules, we can regard that particle to be suspended in a continuous medium of solvent. The particles composing the solvent move because of thermal excitation and collide with the other particles. As a consequence the particles describe a random walk that is called diffusive Brownian motion. The random walk is described by the diffusion coefficient, defined as $D = \frac{\langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle}{6t}$. The transition probability P becomes Gaussian in the limit of small unit step or large number of steps, and can be written as:

$$P(\mathbf{r},\mathbf{r}';t) = (4\pi Dt)^{-3/2} exp\left(-\frac{(\mathbf{r} - \mathbf{r}')^2}{4Dt}\right) \tag{2.24}$$

This probability satisfies the equation of diffusion,

$$\frac{\partial P_x}{\partial t} = D\frac{\partial^2 P_x}{\partial x^2} \tag{2.25}$$

The physical meaning of the probability P can be better understood if we imagine to concentrate all the particles of the sample solution in a position **r** at time 0. The concentration profile is then c(**r**,0). If the particles are free to diffuse for a time interval t they will move to other positions so the concentration profile at a position **r**′ after time t will be the product of the initial concentration and the probability that the particles would change their positions from the starting point to another. $c(\mathbf{r},t) = \int_V P(\mathbf{r},\mathbf{r}';t)c(\mathbf{r}',0)d\mathbf{r}'$

If the particles move according to the diffusion equation, their dynamic structure factor is

$$S_1(\mathbf{k},t) = \int exp[i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')](4\pi Dt)^{-3/2}exp\left(-\frac{(\mathbf{r}-\mathbf{r}')^2}{4Dt}\right)d\mathbf{r} = exp(-D\mathbf{k}^2t) \tag{2.26}$$

In a system of scatterers of equal cross-section and at moderate concentrations (such that multiple scattering can be neglected), it can bee seen that $S_1(\mathbf{k},t) = |g_1(t)| = exp(-\Omega t)$ where the decay rate $\Omega$ is related to the scattering angle as $\Omega = D\mathbf{k}^2$.

## 2.5 Diffusion of polymers in solution

If we consider a solution of polymers instead of particles, the dynamic structure factor becomes

$$S(\mathbf{k},t) = \frac{1}{N_pN}\sum_{m,n=1}^{N_p}\sum_{i,j=1}^{N} < exp[i\mathbf{k}\cdot(\mathbf{r}_{mi}(0)-\mathbf{r}_{nj}(t))] > \tag{2.27}$$

where the coordinates $\mathbf{r}_i$ refer to the positions of the center of mass of the monomers, $N_p$ is the number of chains in the scattering volume and N is the number of monomers for each chain. Also in this case $S(\mathbf{k},t)$ can be decomposed in two parts, the first that is the single chain dynamic structure factor $S_1(\mathbf{k},t)$ and the other that depends on the interactions between different chains.

$$S(\mathbf{k},t) = S_1(\mathbf{k},t) + \frac{N_p}{N}\sum_{i,j=1}^{N} < exp[i\mathbf{k}\cdot(\mathbf{r}_{1i}(0)-\mathbf{r}_{2j}(t))] \tag{2.28}$$

$$S_1(\mathbf{k},t) = \frac{1}{N}\sum_{i,j=1}^{N} < exp[i\mathbf{k}\cdot(\mathbf{r}_{1i}(0)-\mathbf{r}_{1j}(t))] \tag{2.29}$$

In short time scales $S_1(\mathbf{k},\tau)$ exhibits a complicated pattern reflecting complex motions of different parts of the polymer chain. Over a long time, however, the motion is dominated by the translation of the chain as a whole in the solution averaged over all conformations. In 2.30, the displacement between monomer i at time 0 and monomer j at time $\tau$ on chain 1 consists of three parts:

$$\mathbf{r}_{1i}(0)-\mathbf{r}_{1j}(t) = [\mathbf{r}_{1i}(0)-\mathbf{r}_{1G}(0)] + [\mathbf{r}_{1G}(0)-\mathbf{r}_{1G}(t)] + [\mathbf{r}_{1G}(t)-\mathbf{r}_{1j}(t)] \tag{2.30}$$

where $\mathbf{r}_{1G}(0)$ and $\mathbf{r}_{1G}(t)$ are the center of mass positions of the chain at time 0 and $t$, respectively. Initially, the three parts are correlated. As $t$ increases, the chain conformation becomes randomized, and the three vectors become more independent from the others. Thus the statistical average of the square is

$$< [\mathbf{r}_{1i}(0)-\mathbf{r}_{1j}(t)]^2 >=< [\mathbf{r}_{1i}(0)-\mathbf{r}_{1G}(0)]^2 > + < [\mathbf{r}_{1G}(0)-\mathbf{r}_{1G}(t)]^2 > + < [\mathbf{r}_{1G}(t)-\mathbf{r}_{1j}(t)]^2 > \tag{2.31}$$

The first and third terms are equal to the square of the gyration radius by definition. Only the second term grows with time because of diffusion of the chain as a whole. After a long time, the second term becomes dominant. Thus, in the limit of long time

$$6Dt = -6\frac{ln|g_1(\tau)|}{\mathbf{k}^2} =< [\mathbf{r}_{1i}(0) - \mathbf{r}_{1j}(\tau)]^2 >\rightarrow< [\mathbf{r}_{1G}(0) - \mathbf{r}_{1G}(\tau)]^2 > \tag{2.32}$$

and therefore the diffusion coefficient is related only to the motion of the center of mass of the polymers and $g_1(\tau)$ gives information about the center-of-mass diffusion.

### 2.5.1 Hydrodynamic radius of a polymer chain

To drag a particle suspended in a viscous medium at a constant velocity v, a constant force of $F = \zeta v$ must be applied to the particle. The coefficient $\zeta$ is called the friction coefficient. Einstein showed that the diffusion coefficient D of the particle in a quiescent solution at temperature T is related to $\zeta$ by

$$D = \frac{k_B T}{\zeta} \tag{2.33}$$

Stokes showed that the friction coefficient for a sphere of radius $R_S$ is given by

$$\zeta = 6\pi\eta_S R_S \tag{2.34}$$

Combining the two equations above gives the Stokes-Einstein equation, where $R_S$ is the Stokes radius:

$$R_S = \frac{k_B T}{6\pi\eta_s D} \tag{2.35}$$

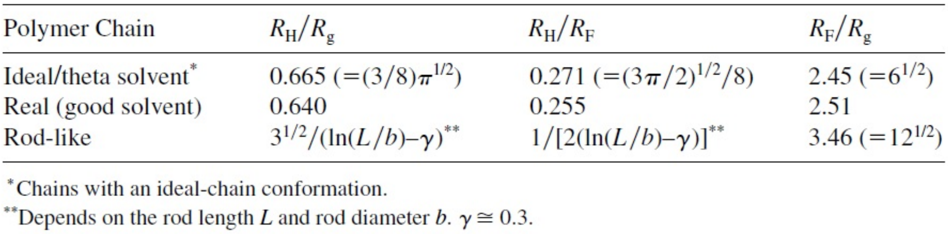
We can extend the concept of the Stokes radius to nonspherical suspensions and molecules. Once the center-of-mass diffusion coefficient D is measured for the suspension or the molecule, we can introduce the hydrodynamic radius $R_H$ by

$$R_H = \frac{k_B T}{6\pi\eta_s D} \tag{2.36}$$

The friction of the chain molecule is smaller than the friction that nonbonded, independently moving N monomers receive. In fact the motion of one of the monomers accompanies motions of adjacent solvent molecules in the same direction, and their effect propagates to another monomer to facilitate its motion in the same direction in an otherwise stagnant solvent. This interaction is called hydrodynamic interaction. Oseen found that the magnitude of the hydrodynamic interaction between two particles at **r** and **r′** is proportional to $|\mathbf{r} - \mathbf{r}'|$. The interaction decays only algebraically with a small exponent of -1 and therefore is long ranged. In a chain molecule, all monomers affect all other monomers because they are close to each other. The hydrodynamic radius of a chain molecule is defined as [14]:

$$\frac{1}{R_H} = \left\langle \frac{1}{|r_m - r_n|} \right\rangle \tag{2.37}$$

The average is taken with respect to possible positions of the two monomers m and n ($m \neq n$) and then with respect to m and n that run over all monomers of the chain. For a Gaussian chain $\frac{1}{R_H} = 8(\frac{2}{3\pi})^{1/2}\frac{1}{b\sqrt{N}}$. For this model the hydrodynamic radius is smaller than the gyration radius and the end to end distance (see Fig.2.5). The gyration radius $R_G$ and the end to end distance $R_F$ are defined as in the equations 2.6 and 2.8.



| Polymer Chain | $R_H/R_g$ | $R_H/R_F$ | $R_F/R_g$ |
|---|---|---|---|
| Ideal/theta solvent* | 0.665 (=$(3/8)\pi^{1/2}$) | 0.271 (=$(3\pi/2)^{1/2}/8$) | 2.45 (=$6^{1/2}$) |
| Real (good solvent) | 0.640 | 0.255 | 2.51 |
| Rod-like | $3^{1/2}/(\ln(L/b)-\gamma)$** | $1/[2(\ln(L/b)-\gamma)]$** | 3.46 (=$12^{1/2}$) |

*Chains with an ideal-chain conformation.

**Depends on the rod length $L$ and rod diameter $b$. $\gamma \cong 0.3$.

**Figure 2.5:** Comparison between hydrodynamic and gyration radii and the end to end distance of a polymer chain for different models.

The comparison between hydrodynamic and gyration radii of a polymer has also been exploited from the experimental point of view, for example through the study of the coil to globule transition of a linear chain of PNIPAM with DLS measurements [15].

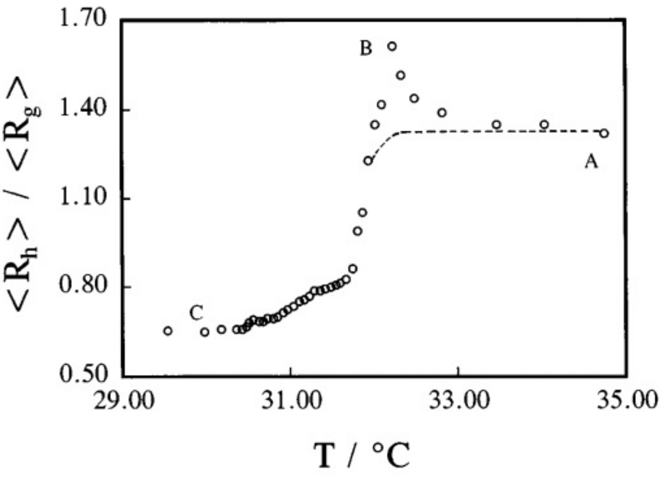


**Figure 2.6:** Ratio between the hydrodynamic to the gyration radius of a linear chain of PNIPAM polymer from [15] at different temperatures

In Fig.2.6 the ratio between the hydrodynamic and gyration radii is plotted as a function of the temperature. To a change of the temperature corresponds a change in the conformation of the polymer. Depending on the conformation the ratio assumes different values: at lower temperatures the polymer is similar to a swollen coil and the hydrodynamic ratio is smaller than the gyration radius. At higher temperatures the conformation changes into a globule and the hydrodynamic radius becomes bigger than the gyration radius. The value of the ratio at the coil conformation (about 0,66) is similar to the ratio calculated for an ideal chain in good solvent.

### 2.5.2 Internal motions of a polymer chain in solution

In a DLS experiment on a solution of polymers in a dilute regime, if we collect the scattered light at angles that satisfy $kR_G < 1$ or if we look at the short timescales of $g_1(\mathbf{k}, t)$ we can study the internal motions of a polymer chain. On the other hand, if the angle is big enough that $kR_G > 1$ or if we look at longer timescales of $g_1(\mathbf{k}, t)$, we see the diffusion of the center of mass of the polymer. The internal motions of a polymer chain were formalized through two main models, the Rouse and the Zimm chains [16].

The Rouse chain was the first successful molecular model; the chain is represented as N beads connected by springs of root mean square size b. The beads in the model only interact with each other through the connecting spring. Each bead is characterized by its own independent friction and the solvent is assumed to be freely draining through the chain. The longer characteristic time of the chain is the Rouse time $\tau_R$, that is the time during which the whole polymer diffuses a distance of the order of its size.

$$\tau_R \approx \frac{R^2}{D_R} \approx \frac{NR^2\zeta}{kT} \tag{2.38}$$

where $D_R$ is the diffusion coefficient of the whole chain and $\zeta$ is the total friction coefficient. On timescales shorter that $\tau_R$ the chain exhibits viscoelastic modes because of the complexity of the interactions between the beads, while on timescales longer the motion is simply diffusive.

### 2.5.3 Rouse relaxation modes

The Rouse time is not the only relaxation time of a Rouse chain. If we consider a natural number p, it can be demonstrated that every section of N/p monomers relaxes with a different timescale.

$$\tau_p \approx \tau_0 \left(\frac{N}{p}\right)^2 \tag{2.39}$$

for p= 1,2,.., N. where $\tau_0$ is the characteristic timescale called the Kuhn monomer relaxation time during which monomer would diffuse a distance of its own size b as it was not attached to the chain. For $p = 1$ we find the Rouse time, while for $p = 2$ we consider the motion of half of the chain, and so on. For big p the motion is faster because a small number of monomers is involved, while for small p the number of monomers grows up and the relaxation becomes slower. At time $\tau_p$ all modes with index higher than p have already relaxed, but modes with index lower than p have not yet relaxed.

The motion for timescales shorter than $\tau_p$ is subdiffusive. In $\tau_p$ the chain section containing $N/p$ monomers diffuse a distance of the order of their size, $b(N/p)^{1/2}$. The mean square displacement is then

$$< [\mathbf{r}_i(\tau_p) - \mathbf{r}_j(0)]^2 > \approx b^2\frac{N}{p} \approx b^2 \left(\frac{\tau_p}{\tau_0}\right)^{1/2} \tag{2.40}$$

For timescales $\tau_0 < t < \tau_p$ the root mean square is

$$< [\mathbf{r}_i(t) - \mathbf{r}_j(0)]^2 > \approx b^2\frac{N}{p} \approx b^2 \left(\frac{t}{\tau_0}\right)^{1/2} \tag{2.41}$$

For the motion to be diffusive, the mean square displacement must be linear in time, so the diffusion coefficient must depend on time.

$$< [\mathbf{r}_i(t) - \mathbf{r}_j(0)]^2 > \approx D(t)t \approx t^{1/2} \tag{2.42}$$

### 2.5.4 Zimm relaxation modes

The Rouse model ignores hydrodynamic interaction forces, and assumes that the beads only interact through the springs that connect them. In dilute solutions, hydrodynamic interactions between the monomers in the polymer chain are strong. These hydrodynamic interactions are strong also between the monomers and the solvent within the pervaded volume of the chain. When the polymer moves, it effectively drags the solvent within its pervaded volume with it. The Zimm model treats the pervaded volume of the chain as a solid object moving through the surrounding solvent. This model predicts that a chain diffuses as a particle with volume proportional to the chain's pervaded volume in solution. In this case the diffusion coefficient of a chain becomes

$$D_Z = \frac{8}{3\sqrt{6\pi^3}} \frac{kT}{\eta_s R} \tag{2.43}$$

and the time during which the chain diffuses a distance of its own size is $\tau_Z$

$$\tau_Z \approx \frac{R^2}{D_Z} \approx \tau_0 N^{3\nu} \tag{2.44}$$

In dilute solution the Zimm time in shorter than the Rouse time. Zimm motion has less frictional resistance than Rouse motion, and therefore, the faster process is Zimm motion.

Zimm model also predicts a chain to have internal relaxation modes. There are a number of modes equal to the number of monomers also for this model. The relaxation time of the pth mode is

$$\tau_p \approx \tau_0 \left(\frac{N}{p}\right)^{3\nu} \tag{2.45}$$

As in the Rouse model, the mean-square displacement of monomer j during time $\tau_p$ is of the order of the mean-square size of the section containing N/p monomers involved in a coherent motion at this time:

$$< [\mathbf{r}_i(\tau_p) - \mathbf{r}_j(0)]^2 > \approx b^2 \left(\frac{N}{p}\right)^{2\nu} \approx b^2 \left(\frac{\tau_p}{\tau_0}\right)^{2/3} \tag{2.46}$$

The motion of the chain for time $\tau_0 < t < \tau_p$ is subdiffusive because the root mean square deviation is proportional to $t^{2/3}$

$$< [\mathbf{r}_i(\tau_p) - \mathbf{r}_j(0)]^2 > \approx b^2 \left(\frac{t}{\tau_0}\right)^{2/3} \tag{2.47}$$

Zimm and Rouse models are two extremely different ways of representing a polymer chain in a diluted regime. The real motion of the chain stands in between the two models.

### 2.5.5 Particle sizing method

From the correlation function of the electric field it is possible to obtain the hydrodynamic radius of a particle in solution in a dilute regime by its relation with the diffusion coefficient. If we measure a solution of polymers the situation is more complicated because at a fixed temperature a polymer chain does not have a unique conformation, it can assume many conformations, each characterized by probability. Different conformations mean different hydrodynamic radii, and different diffusion coefficients. The correlation function becomes a superposition of exponential decays, one for every possible conformation of the chain. In a dilute solution where the scattered intensity comes from the superposition of the scattered radiation from each polymer in the scattering volume, there is the superposition of the scattered intensity coming from each population of polymers with the same dimension. DLS is not a single molecule technique, so it is not possible to separate the scattering contribution coming from molecules with a very similar dimension and molecular weight.

$$g_1(t) = \int G(\Omega) e^{-\Omega t} d\Omega \tag{2.48}$$

where $G(\Omega)$ represents the contribution of each population. The integral in 2.48 can be written as a stretching exponential, where $\alpha$ is the stretching exponent:

$$g_1(t) = A e^{-(\frac{t}{\tau})^{\alpha}} \tag{2.49}$$

The average relaxation time becomes

$$< \tau > = \int_0^{\infty} e^{-(\frac{t}{\tau})^{\alpha}} dt = \tau \frac{1}{\alpha} \Gamma \left( \frac{1}{\alpha} \right) \tag{2.50}$$

From the average relaxation time the average hydrodynamic radius is determined:

$$R_H = \frac{< \tau > k_B T k^2}{6 \pi \eta} \tag{2.51}$$

## 2.6 DLS instrumentation

A scheme of the geometry of the DLS apparatus (ST100, Scitech Instruments) used during this work is shown in Fig.2.7.

A cylindrical glass tube containing a sample solution is immersed in a chamber full of silicone oil with the refractive index (n=1,39) similar to the one of the glass. The oil is thermostatted and flown inside the chamber by a thermal cycler (Julabo F25-HE). The incoming beam comes from a laser (solid-state, Nd-YAG) at a wavelength of 532 nm and is brought inside the optical chamber through an single-mode optical fiber. The scattered light is collected at 90° from the direction of the incident light by another fiber that is connected to two photomultipliers. There are two PMTs in order to capture the signals with a timescale smaller than the dead time of the PMT that is approximately about $10^{-7} s$. The signal then is sent to a correlator (Digital Correlator Flex03LQ-1) that calculates the correlation function of the scattered light.

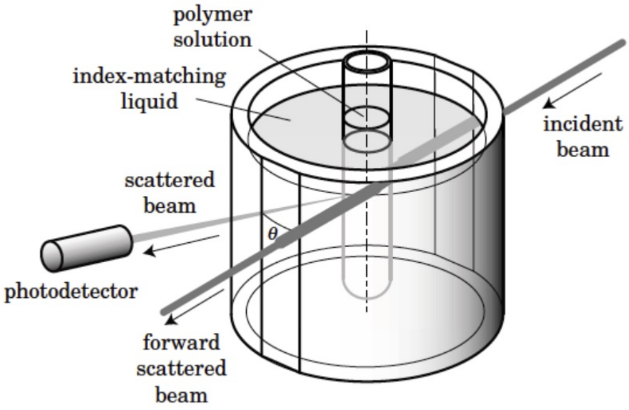


**Figure 2.7:** Scheme of a DLS apparatus.

### 2.6.1 Sample preparation for DLS

The scattering cell is a cylindrical capillary made by borosilicate optical glass (from Hilgenberg GmbH) with an internal diameter of 2,4 mm and an external diameter of 3 mm. The internal part of the cell was typically cleaned with about 20 ml of filtrated acetone, ethanol, 2-propanole and milliq water before the measurements. 2-propanole was also used to clean the external surface.

The concentration of the ssDNA in the sample (0,3 g/l, i.e. 13 nM) was chosen to have enough scattering signal but also to be smaller (about 1/30) of the entanglement concentration so that the experiments were made in a dilute condition. We dissolved the linear M13mp18 in a solution of deionized (milliq) water at two ionic strengths, 10 and 50 mM NaCl.

The scattered light was collected at 90° from the position of the incident light. The temperature varied from 30°C to 5°C with a step of about 2°C. The thermalization time was 40 minutes for each temperature. For each measurement the scattered light was collected for 900 s, then correlated.

## 2.7 Fluorescence Correlation Spectroscopy

Fluorescence Correlation Spectroscopy (FCS) is a technique that measures the diffusion of fluorescent molecules dissolved in a solution in a dilute condition. The basic apparatus is coupled with a confocal microscope, where the sample cuvette is positioned on the stage. A laser beam is focused inside the sample in such a way that the beam waist is smaller than the dimension of the cuvette. The intersection between the excitation volume and the detection volume is called the confocal volume. When a molecule goes through the confocal volume it is excited by the laser and as a consequence it emits fluorescence light, that is collected by a single photon sensitive detector. As long as the molecule remains into the confocal volume the detector records the fluorescence signal. If the solution is sufficiently diluted and there are no free fluorophores it is possible to detect the average number of fluorescent particles inside the confocal volume and the average diffusion time through the confocal volume. FCS is a technique that allows to detect the fluctuations of the number of molecules in the confocal volume and can be used with very diluted systems (a few molecules in a volume of about $1\,\mu m^3$, that corresponds to concentrations of the order fo $1$ to $10\,nM$).

To measure these quantities, the fluorescence correlation intensity is calculated. It shows a characteristic shape like an exponential decay. The characteristic time of the decay is the average permanence time of a molecule inside the confocal volume. If a molecule is close to the center of the detection volume, there will be a high probability to detect a large number of consecutive fluorescence photons from this molecule, i.e. the fluorescence signal will be highly correlated in time. When the molecule (due to diffusion) starts to exit the detection volume, this correlation will continually decrease in time, until the molecule has completely diffused out of the detection volume and the correlation is completely lost (correlation curve drops to zero). The faster a molecule diffuses, the faster the correlation will be lost. If the dimension of the confocal volume is known, it is possible to calculate the diffusion coefficient of the molecule and then the hydrodynamic radius. However it is rather difficult to know that volume because of the complexity of the apparatus, so a technique (Dual Focus Fluorescence Correlation Spectroscopy, 2fFCS) was developed to overcome this problem.

2fFCS employs two laser sources in order to generate two laterally shifted but overlapping laser foci at a fixed and known distance inside the sample. Thus, for each laser focus the corresponding intensity autocorrelation function can be calculated. However, in addition to the autocorrelation curves, one can also correlate the fluorescence signal from one laser focus with that of the second focus. By doing so, a cross-correlation curve is generated which is shifted to longer lag times and has a lower amplitude with respect to the auto-correlation curves. The shape of an autocorrelation curve is completely determined by the detection volume, whereas the shape of the cross-correlation curve is also dependent on the dimension of the overlap region of the two foci. If the distance between both lasers is known, a global fitting of both auto- and cross-correlation curves can yield the absolute diffusion coefficient. This is because the relative time shift between the cross correlation and auto correlation curves scales with the square of the foci distance divided by the diffusion coefficient.

2fFCS has proved to be a very accurate and robust method of measuring diffusion of molecules in dilute regime (from 0,1 up to about 5 nM) [17][18].

Each 2fFCS measurement output consists of three correlation functions. The analysis is made as follows. Both the cross- and auto-correlation functions depend on the shape and the size of the confocal volume, that is described by the molecule detection function $U(\mathbf{r})$. This function calculates the position-dependent efficiency to excite and detect a fluorescence photon from a single molecule. $U(\mathbf{r})$ is successfully represented as the product of three functions, as shown in 2.52

$$U(\mathbf{r}) = \frac{k(z)}{\omega^2(z)} exp\left[-\frac{2(x^2+y^2)}{\omega^2(z)}\right] \tag{2.52}$$

where x, y, and z are cartesian coordinates with the z-axis along the optical axis, and the functions $\omega(z)$ and $k(z)$ are given by

$$\omega(z) = \omega_0\sqrt{1+\left(\frac{\lambda_{ex} z}{\pi\omega_0^2 n}\right)^2} \tag{2.53}$$

and

$$k(z) = 1 - exp\left(-\frac{2a^2}{R_0^2 + (\lambda_{em} z/\pi R_0 n)^2}\right) \tag{2.54}$$

where $\lambda_{ex}$ and $\lambda_{em}$ are the excitation and center emission wavelengths, respectively, $n$ is the sample refractive index, $a$ is the confocal pinhole radius and $R_0$ is a free fit parameter. $\omega(z)$ represents the scalar approximation for the radius of a diverging laser beam with beam waist $\omega_0$, and $k(z)$ depends on the point spread function of confocal imaging. [17].

The cross correlation intensity is

$$\begin{aligned} g_2(t) =& g_\infty(\delta) + 2\epsilon_1\epsilon_2 c\sqrt{\frac{\pi}{Dt}} \times \int_{-\infty}^{\infty} dz_1 \int_{-\infty}^{\infty} dz_2 \frac{k(z_1)k(z_2)}{8Dt + \omega^2(z_1) + \omega^2(z_2)} \\ & \times exp\left[-\frac{(z_1 - z_2)^2}{4Dt} - \frac{2\delta^2}{8Dt + \omega^2(z_1) + \omega^2(z_2)}\right] \end{aligned} \tag{2.55}$$

where $\delta$ is the lateral distance between the foci, $\epsilon_1$ and $\epsilon_2$ are two factors proportional to the overall excitation intensity and detection efficiency in each laser, c is the concentration of the fluorescent molecules, and D is the diffusion coefficient. For calculating the fitting function of the intensity auto-correlation of each focus the formula is the same as for the cross-correlation. We have just to put $\delta = 0$ and replace $\epsilon_1\epsilon_2$ by either $\epsilon_1^2$ or $\epsilon_2^2$ respectively. When fitting the experimentally measured data, one fits the two intensity auto-correlations (which are identical in shape) and the cross-correlation simultaneously, having as fit parameters $\epsilon_1\sqrt{c}$, $\epsilon_2\sqrt{c}$, $\omega_0$, $R_0$ and D. e. The important parameter determining the absolute accuracy of the resulting values of the diffusion coefficient is the lateral distance $\delta$ between the foci. Once fitted the intensity correlations gives the diffusion coefficient and thus the hydrodynamic radius of the molecules observed.

## 2.8 FCS apparatus

We used a 2fFCS apparatus to measure the hydrodynamic radius of M13mp18 as a function of temperature. A scheme of such apparatus is shown in Fig.2.8.

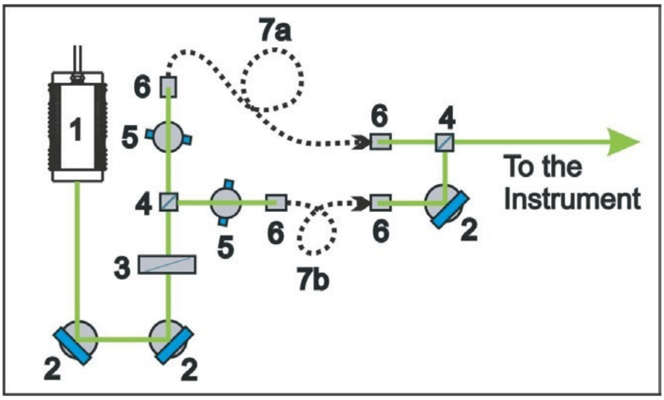


**Figure 2.8:** Sketch of a single laser set-up for dual-focus experiments, equipped with: 1) laser head, 2) mirror, 3) adjustable zeroth-order halfwave plate, 4) polarizing cube, 5) beam displacer, 6) fiber coupler, 7a) long single-mode fiber, 7b) short single-mode fiber. [17]

The specifics The set-up used is based on a PicoQuant MicroTime 200 confocal microscope. To generate the two shifted foci, two orthogonally polarized, pulsed diode lasers (PicoQuant LDH-P-C-640B) are combined with a polarization sensitive beamsplitter and afterwards coupled into a polarization maintaining single-mode fiber. At the fiber output the light is again collimated by an appropriate lens. Both lasers are pulsed alternately. The combined light consists of a train of laser pulses with alternating orthogonal polarization. Before the objective (immersion in water) the laser beam passes through a Nomarski prism (Olympus U-DICTHC) that is normally exploited for differential interference contrast (DIC) microscopy. The principal axes of the Nomarski Prism are aligned with the orthogonal polarizations of the laser pulses, so that the prism deflects the laser pulses in two different directions according to their corresponding polarization. After focusing the light through the objective (Olympus UPLANAPO, 60x, water immersion, NA 1.2), two overlapping excitation foci are generated with a small lateral shift between them. The distance between the beams for a given wavelength is defined by the DIC prism and the objective. As long as these two elements are used, the distance will remain accurately constant over time. Fluorescence is collected by the same objective (epi-fluorescence set-up), passed through the DIC prism and the dichroic beamsplitter and is focused into a single circular aperture (confocal pinhole). Behind the pinhole the light is collimated and divided by a non-polarizing beamsplitter cube and focused onto two single-photon avalanche diodes (SPAD, PDM series, Micro Photon Devices). When calculating correlation functions, only photons from the two different detectors are correlated to prevent distortions of the resulting auto-correlation functions by SPAD after-pulsing. A Time-Correlated Single Photon Counting unit (PicoHarp 300) is used for data acquisition, operating in the time-tagged time-resolved mode, which allows to record for every detected photon its arrival time with a temporal resolution in the nanosecond range and its arrival time with respect to the last laser pulse with picosecond timing resolution (TCSPC time). The TCSPC time of each recorded photon is used to determine in which laser focus/detection region the fluorescence light was generated. The photons detected from the single photon counting unit during a detection time were summed. The square of the ratio between the mean of the number of photons detected during the detection time and their variance is the average number of molecules in the confocal volume. At the concentration explored for ssDNA measurements the average number of molecules was about 60.

### 2.8.1 Sample preparation for FCS

The sample for FCS consisted of a solution of 2nM linear M13mp18, deionized (milliq) water at 50mM NaCl. The cell was cylindrical with a volume of about 50 $\mu l$. The solution was vortexed before the measurement to assure the maximum concentration uniformity.

The temperature was varied between 5°C and 30°C with a step of 2°C. Each temperature was sampled twice with correlations of 10 min.

## 2.9 Results

We measured the hydrodynamic radius ($R_H$) of a diluted sample of linear ssDNA dispersed in a solution at different ionic strengths. The measurements were collected at various temperatures within a range from 5°C to 40°C through Dynamic Light Scattering and Fluorescence Correlation Spectroscopy. The results obtained with both techniques show a good compatibility. To understand the thermodynamics responsible for the behaviour of $R_H$ vs temperature we studied the number, the distribution and the free energy loss of the possible intra-polymer interactions.

### 2.9.1 DLS results

Using Dynamic Light Scattering we measured the scattered intensity at 90° from the incident light for samples of linear M13mp18 dissolved in a solution at 10 and 50 mM NaCl. For the solution at 10 mM NaCl we collected the scattered light while decreasing the temperature. For the 50 mM NaCl solution we also collected the scattered light while increasing the temperature, showing a small hysteresis of the behaviour of the chain between the two temperature ramps. At every temperature we collected 4 times the scattered intensity for 900 s, we calculated the 4 correlation functions of the intensity and we averaged them. In Fig.2.9 there are two intensity correlation functions of a sample at 50 mM NaCl at a temperature of 10°C (red line) and 25°C (blue line). They are normalized to understand the two different decay times. We fitted such correlations with a stretched exponential, as explained in section 2.5.5.

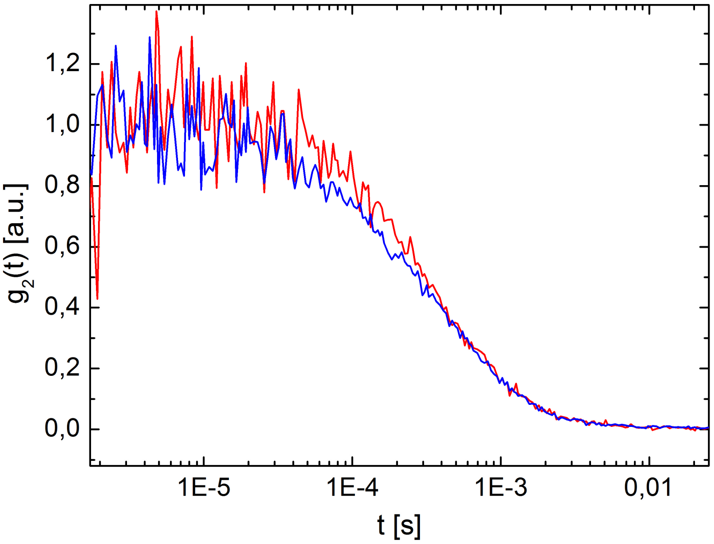


**Figure 2.9:** Intensity correlation function for a sample of M13mp18 at 0,3 g/l dissolved in a solution of 50 mM NaCl at 10°C (red line) and 25°C (blue line). The curves are normalized.

In Fig.2.10 there are some examples of fitted correlation functions at different temperatures for a sample at 50 mM NaCl. The fitting parameters are listed in table 2.1.

We noticed that the decay time of the intensity correlation function increases as temperature is decreased, that may suggest that the dimension of the molecules we are looking at should increase. Nevertheless the hydrodynamic radius decreases as temperature is decreased because the viscosity of the solution (here approximated to the viscosity of water, due to the very dilute concentration of salts and DNA) changes of a factor of 2 in the temperature range of the measurements. To understand if the change in dimension was only a result of the viscosity change and not of the closure of contacts within the chain, we measured the hydrodynamic radius at different temperatures of a non-

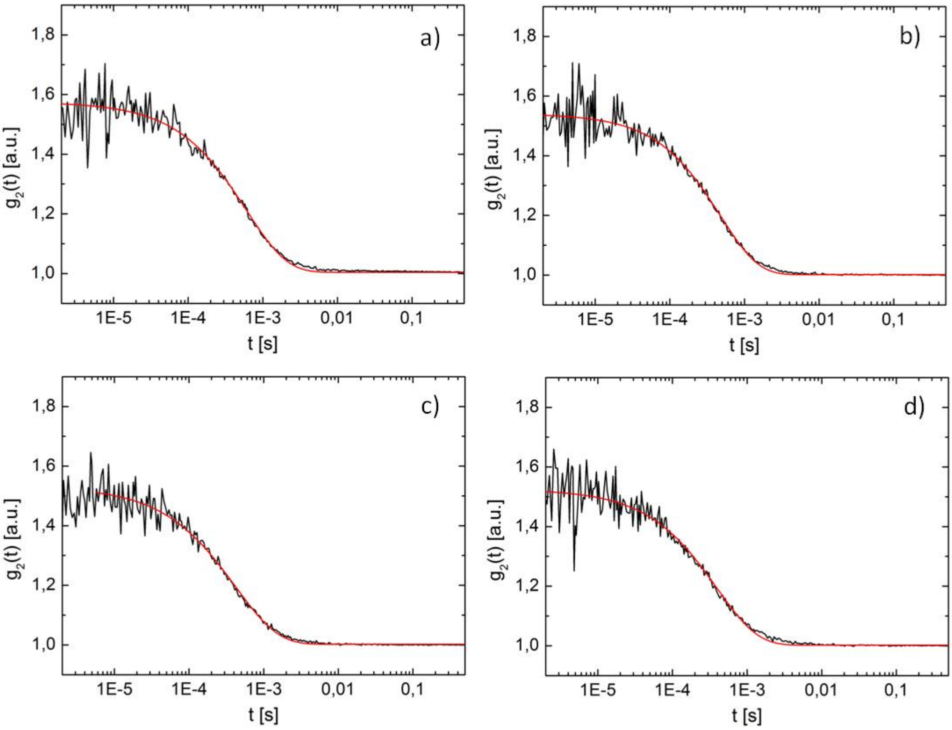


**Figure 2.10:** Examples of fitted intensity correlation functions for a sample of M13mp18 at 0,3 g/l dissolved in a solution of 50 mM NaCl. The correlations were measured this chronological order at a) 5°C, b) 15°C, c) 25°C and d) 30°C.

| Salt concentration [mM] | T [°C] | $<\tau>$ [s] | $\alpha$ | $R_H$ [nm] |
|---|---|---|---|---|
| 50 | 5 | 0.000354 | 0.83 | 23.66 |
| 50 | 15 | 0.000255 | 0.86 | 23.31 |
| 50 | 25 | 0.000246 | 0.79 | 29.49 |
| 50 | 30 | 0.000223 | 0.83 | 29.69 |

**Table 2.1:** Fitting parameters of the intensity correlations plotted in Fig.2.10 measured with DLS as a function of temperature and salt concentration. $R_H$ is the hydrodynamic radius, $\tau$ is the average decay time, calculated as in 2.5.5, and $\alpha$ is the stretching exponent. The measurements were done at a fixed scattering angle of 90° while increasing the temperature.

interacting chain of DNA of a length comparable to M13mp18. The chain considered was poly-A (Sigma-Aldrich, P9403), that has no interactions between the monomers because it is composed only of nucleotides with adenine nucleobases. We prepared a sample at a concentration of 0,3 g/l in a buffer at 50 mM NaCl, we vortexed and heated the solution in order to dissolve the DNA. We filtrated the sample to eliminate any possible residual aggregates, originated from the fact that polyA is sold as a lyophilized powder. In Fig.2.11 the intensity correlation functions collected at different temperatures are plotted. We observed a uniform shift of the correlations, that is due only to the change of viscosity of the solution. We normalized the correlations and we divided the timescale for the water viscosity at the different temperatures sampled: all the curves superimposed (see Fig.2.12) to confirm that the shift seen before was just caused by the change of viscosity. The correlations show the presence of multiple exponential decays because of the high polydispersity of the sample, that has a length between 600 and 4000 nucleotides. The decay times are too long to be related to the motion of a single chain of polyA. There must be some aggregates in the solution that we could not eliminate that affect much the measurements. Nevertheless the role of viscosity in confirmed, even though much study on this side is required.

In Fig.2.13 we plotted the hydrodynamic radius ($R_H$) of the linear M13mp18 as a function of the temperature. The sample was stored at 5°C and then measured at various temperatures, from the lowest (5°C) to the highest (30°C). Even though the temperature range is not well sampled it is possible to tell two regions apart, one at low temperatures (lower than 5°C) where the chain has a dimension of about 20 nm, and one at high temperatures (from 20 to 30°C) where the dimension is close to 35 nm. This behaviour suggests that the chin undergoes a transition from a globular to a coil conformation between 5 and 20°C. In section 2.10 we discuss the behaviour of the transition as a function of the temperature.

In Fig.2.14 we showed the behaviour of the hydrodynamic radius measured while decreasing the temperatures of the sample for a solution of linear M13mp18 at 50 mM NaCl. As for the sample at 10 mM of NaCl, this solution was stored at 5°C before being measured. For temperatures smaller than 15°C the measured $R_H$ is about 24 nm, that increases to about 30 nm in the range of temperatures between 15 and 25°C. At temperatures higher than 25°C the dimension is almost constant and of a value of about 30nm.

We repeated a measurement with the same sample while increasing the temperature and we obtained the hydrodynamic radii shown in Fig.2.15. The coil and globule dimensions are compatible with the ones plotted in Fig.2.14, while the transition occurs in a more narrow temperature range than the measurements with decreasing temperature, approximately from 18 to 22°C.

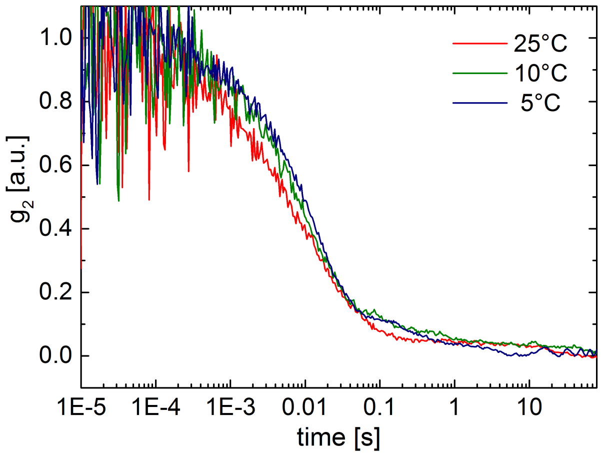


**Figure 2.11:** Intensity correlation functions of poly-A at a DNA concentration of 0,3 g/l and 50 mM NaCl. Temperatures are indicated in the legend.

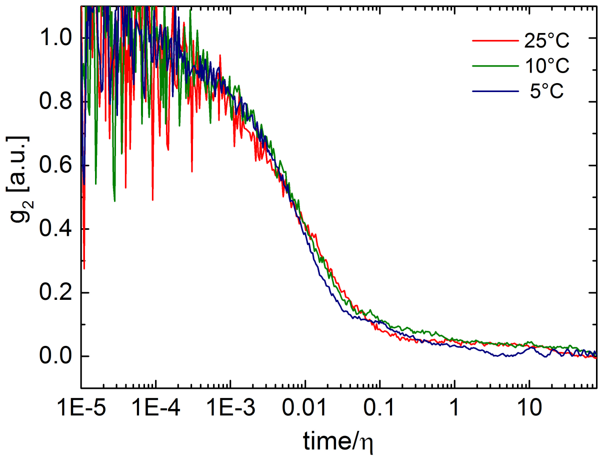


**Figure 2.12:** Normalized intensity correlation functions of poly-A at a DNA concentration of 0,3 g/l and 50 mM NaCl. Time is divided for the water viscosity at the corresponding temperature.

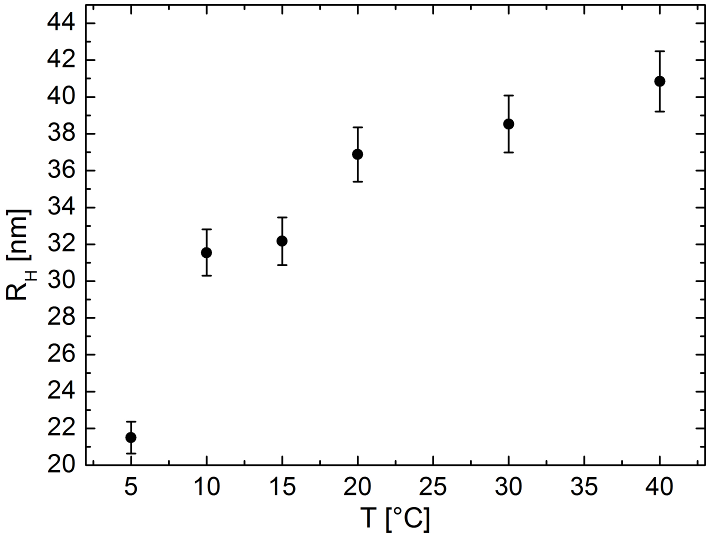


**Figure 2.13:** Hydrodynamic radius of M13mp18 at different temperatures for 10 mM NaCl buffer. The scattering angle was kept fixed at 90° for all measurements.

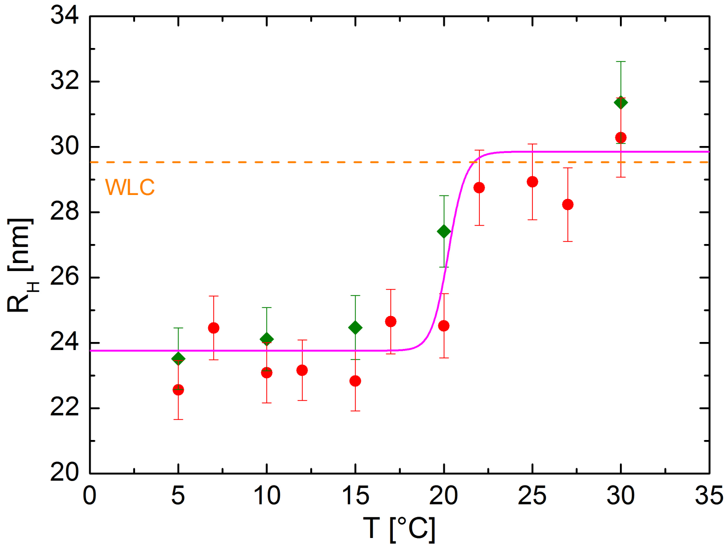


**Figure 2.14:** Hydrodynamic radius of M13mp18 at different temperatures for 50 mM NaCl. The measurements were done while decreasing the temperature. The scattering angle was kept fixed at 90° for all measurements. The magenta line is drawn to guide the interpretation of the data.

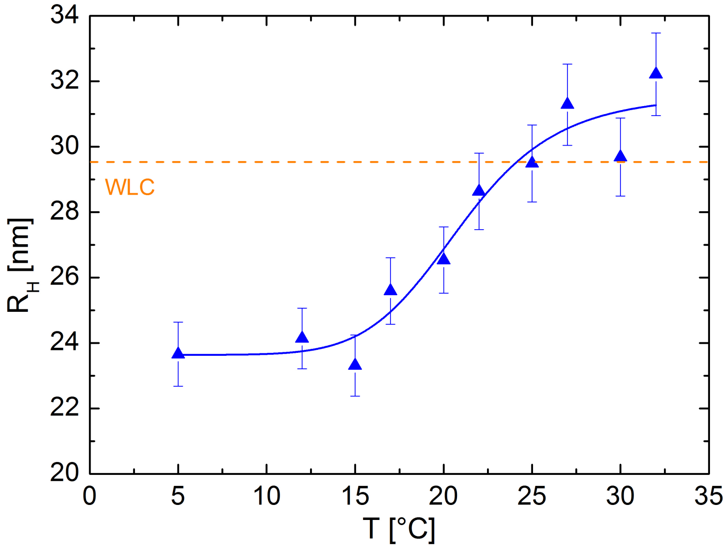


**Figure 2.15:** Hydrodynamic radius of M13mp18 at different temperatures for 50 mM NaCl. The measurements were done while increasing the temperature. The scattering angle was kept fixed at 90° for all measurements.

In Fig.2.16 all the M13mp18 hydrodynamic radii measured with DLS at different temperatures are compared.

The temperature at which the transition occurs is smaller for the 10 mM sample than for the 50 mM. The shift can be explained considering that the melting temperature of a duplex of DNA increases when the salt concentration in the buffer is increased. This happens for all the segments of chains that bind together, so the globular state is stabilized at higher temperatures.

Looking at the measurements performed at the temperature in between 10 and 20°C we can infer that the M13mp18 exhibits a continuous transition between coil and globule. It is difficult to say if the $R_H$ measured between 10 and 20°C represent an effective intermediate conformation between coil and globule or if they are just the average between the two. The correlation functions in this intermediate temperature range do not exhibit two clearly different decays, one for the globule and one for the coil conformation, so we can infer that the decay rate correspond to an intermediate configuration, where some contacts are closed. Moreover the stretching exponent does not increase the temperature transition range, so the distribution of conformations must be equal at all the temperatures.

From Figs.2.14 and 2.15 we noticed some difference between the hydrodynamic radii measured while increasing the temperature and the ones measured decreasing the temperature. It is possible that the kinetics of the collapse and of the swell of the chain are different because the thermal stories of the sample affect the temporal sequence of closure of the contacts. In fact the samples observed while increasing the temperature have different were diluted at room temperature, and suddenly stored at 5°C, so it is

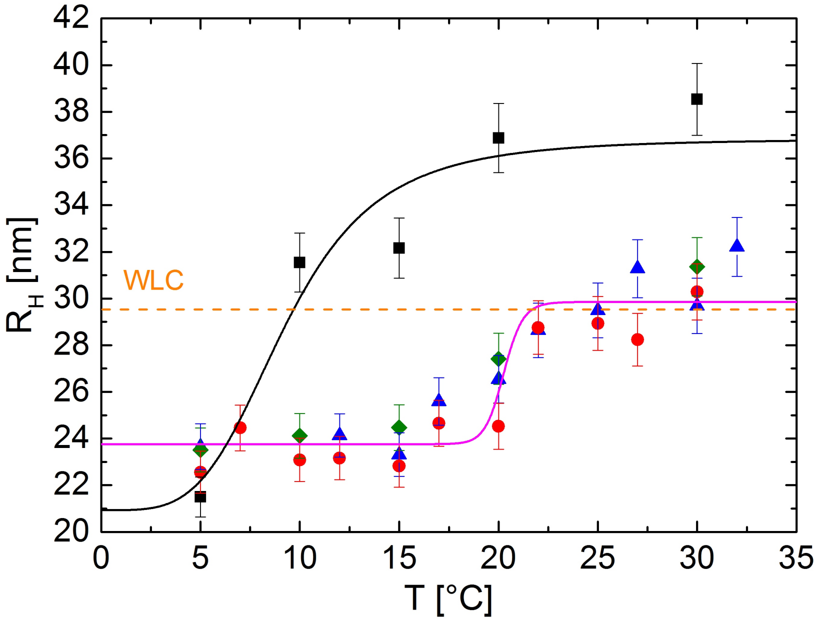


**Figure 2.16:** All the hydrodynamic radii measured with DLS of samples of M13 at different ionic strengths. Black squares refer to the sample at 10 mM of NaCl, red circles and green diamonds are measured from samples at 50 mM NaCl while decreasing the temperatures. The blue triangles come from the samples at 50mM NaCl measured with increasing temperature.

possible that the chain assumed a conformation that is not the minimum of free energy at that temperature. On the opposite, the sample measured from 30 to 5°C had the opportunity to find its minimum of free energy at every temperature because the cooling rate was very small (30 minutes of thermalization at every temperature, 30 minutes of measurements and steps of about 2°C). This aspect should be further investigated.

In a typical scattering experiment of interacting molecules it is possible to observe an aggregation process during the temperature change. At high concentrations aggregation is expected to occur frequently, while at low concentrations it should be a negligible phenomenon. However scattering experiments require a big enough concentration to see some scattering signal from the molecules in a reasonable leg of time, so we had to chose a concentration for the experiments that had to be a compromise between a value big enough to have scattering signal and at the same time small enough to have a diluted solution. To verify whether aggregation happened during our measurements we computed the average scattered intensity for all the temperatures sampled. We found that the average scattered intensity $< I_s >$ didn't change with the temperature for both ionic strengths, to confirm that there is no aggregation process during the measurements. In Fig.2.17 $< I_s >$ is plotted as a function of temperature for the sample at 50 mM NaCl.

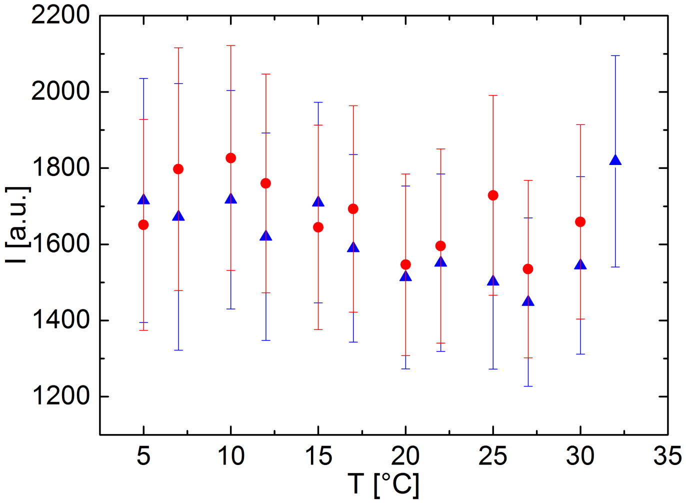


**Figure 2.17:** Average scattered intensity measured with DLS for a sample of 0.3 g/l at 50 mM NaCl at different temperatures. The error bars are the standard deviation of the intensity.

### 2.9.2 FCS results

FCS measurements at a DNA concentration of 2 nM also show a transition between globular and coil conformation.

Each point in Fig.2.18 and 2.19 is averaged over 3 data acquisitions. All the measurements of the intensity over time were analyzed in order to eliminate the occasional signals due to impurities o the sample, that affect all the analysis. To reduce the errors coming from the fitting procedure of the correlation functions, we fitted the value of $\omega_0$

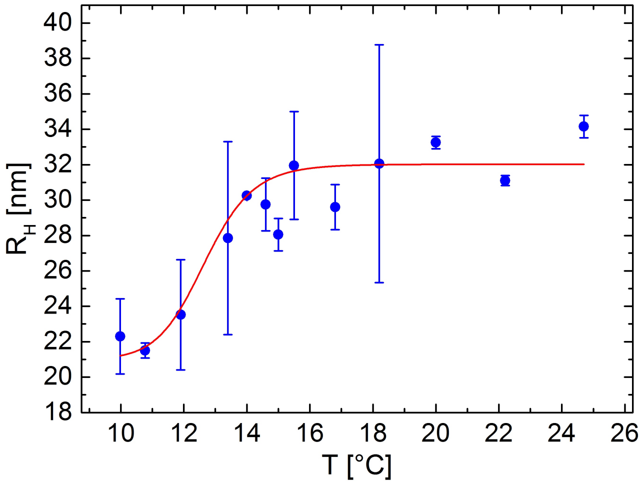


**Figure 2.18:** Hydrodynamic radius of M13mp18 at different temperatures measured with FCS at 50 mM NaCl while increasing the temperature.

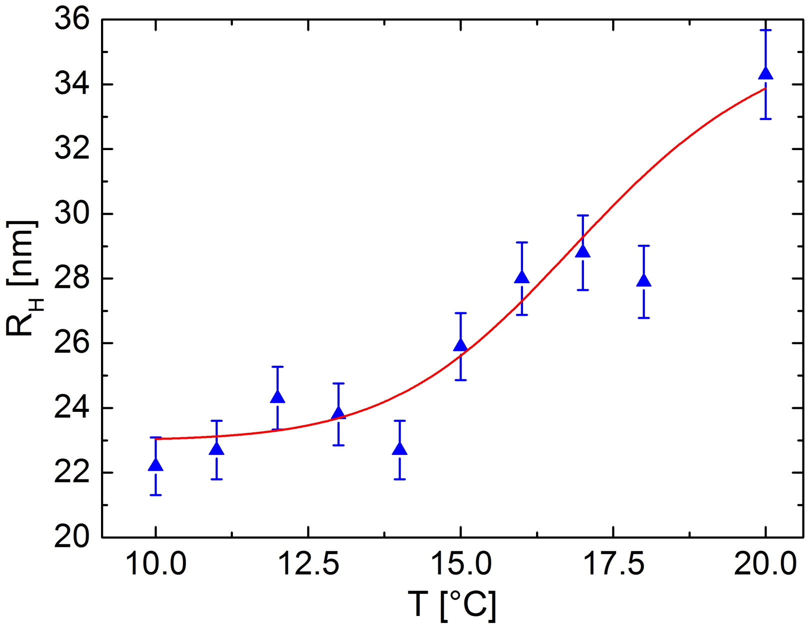


**Figure 2.19:** Average hydrodynamic radius of M13mp18 at different temperatures measured with FCS at 50 mM NaCl while slowly decreasing the temperature.

for both samples because it is reasonable that during the measurements the beam waist would not change if the power of the laser remains the same. The measurements of the first sample were done at a distance of $10\mu m$ from the focal plane of one focus, while for the second the distance was increased up to $30\mu m$ to better avoid temperature gradients.

The samples used for the two curves have a different thermal histories. The first sample (see Fig.2.18) has been stored at -20°C for some hours, thawed to 5°C and then its dimension was measured from 5°C to 30°C with steps of 2°C. The second sample (see Fig.2.19) was never frozen and has been stored at 5°C for some hours before the measurements. Then it was gently brought to 20°C and then down to 10°C. The $R_H$ was measured from 10°C to 25°C at different intermediate steps. There are some differences between the two measurements that can be accounted for the thermal history of the samples. The first sample has a bigger globule dimension than the second, probably because the freezing procedure had caused the chain to suddenly fold into a globular configuration that was not the minimum of energy, while for the second sample the cooling process happened to be gentle and the chain could rearrange in a more favourable configuration. As a consequence the temperature at which the transition occurs is different for the two samples: the second sample has a higher transition temperature than the first sample because the cooling procedure had allowed it to close more contacts and/or the most energetic ones.

The second (Fig.2.19) FCS measurements confirm the DLS results: the hydrodynamic radius of both coil and globule are similar to the one measured with DLS at the same ionic strength. Also the transition between the two conformations occurs in the same range of temperatures.

The intensity correlations measured in the range from 10 to 15°C seem to have just one decay that can be attributed to the ssDNA, to prove that the change from coil to globular conformation happens gradually. In Fig.2.20 there are the intensity correlations calculated from an acquisition at NN°C. The light orange and the blue curves are the auto-correlations measured at the two foci, while the dark orange curve is the cross-correlation between the two foci.

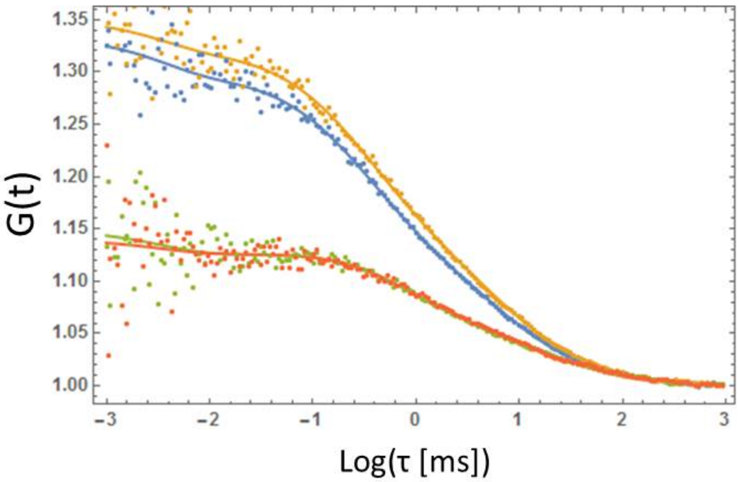


**Figure 2.20:** Intensity correlation functions measured for the first sample at 15°C and 50 mM NaCl.

The measurements with the first sample were repeated while increasing and decreasing the temperature, but the results were not the same. While decreasing the temperature the $R_H$ did non decrease and kept approximately the same dimension as for the coil conformation. This behaviour could be due to a difference in the kinetics of the sample during the cooling and heating processes. We couldn't investigate further the phenomenon, but specific measurements should be done.

The FCS measurements were strongly affected by the impurities inside the sample. The purification of the linear ssDNA had an efficiency of about the 50%: almost half of the primers used for the cut could not be eliminated. The primers used are bound with the fluorophore so the FCS apparatus detects them and all the correlation functions showed two decays, one with the shortest characteristic time that well describes the theoretical motion of the primer and one of M13mp18. The presence of a population of fluorescent primers adds an extra decaying function in the correlation fitting function thus introducing one extra parameter. We measured the diffusion coefficient of a sample of primers in solution (without M13mp18) and we constrained the fastest exponential of the correlation function of the sample of M13mp18 to have this diffusion coefficient, in such a way to eliminate the extra parameter from the fitting procedure.

## 2.10 Discussion

From DLS and FCS measurements we obtained some information about the dimension in the coil state of the ssDNA chain studied, that is reasonable if compared to the standard polymer models. We also noticed that the dimension of the chain has a dependence on the temperature. At low temperatures (the range depends on the ionic strength of the solution) the chain is more compact while at high temperatures (approximately above 25°C) the dimension of the chain grows. The values of the dimension in the two conformations are still uncertain because the results obtained with the different techniques used are similar but not compatible. Also the kinetics of the measurements seem to affect the dimensions. However we found a variation of the dimension of the chain (from 25 to 34 nm or from 22 to 30 using FCS at 50 mM NaCl, or from 22 to 30 through DLS at 50 mM and from 22 to 36 with 10 mM NaCl) while changing the temperature. The range of temperatures at which the transition occurs (from 15 to 25°C at 50 mM NaCl or form 5 to 20°C at 10 mM NaCl) was measured with both techniques.

DLS and FCS measurements show that decreasing the temperature from 30 to 5°C the ssDNA changes its conformation from a swollen state to a more compact one. The process must be a consequence of the intra-chain interactions that become stronger at low temperatures and at higher ionic strength. Every time a contact closes a loop forms in the chain because the contact involve only regions of the same chain. As shown before it is possible to find all the interactions that may happen inside the chain, so we tried to study the thermodynamics of these interactions to figure out the sequence of contacts closure of the ssDNA while decreasing the temperature.

Each contact is described by the melting temperature ($T_m$), which is an indicator of the strength of the contact. At temperatures lower than $T_m$ the contact is closed, while at higher it opens. There is an analytic model that allows to calculate the melting temperature of hairpins with a loop of n nucleotides if the entropy and enthalpy variation for a loop of $x$ bases are known [19], as shown in 2.56.

$$Tm = \frac{\Delta H_0}{\Delta S} = \frac{\Delta H_x}{\Delta S_x + C_1 ln(\frac{n}{x})} \tag{2.56}$$

To use 2.56 we have to know the values of enthalpy ($\Delta H_x$) and entropy ($\Delta S_x$) for a hairpin of a loop length of x bases at the experimental ionic strength [19], that account for the number and kind of bases involved in the hairpin, the DNA and the salt concentrations. $C_1 ln(\frac{n}{x})$ is the entropy loss due to the formation of the loop of length $n > x$. This calculation was validated by experimental measurements for a loop length up to 35 bases [19]. For M13mp18 the possible loops are much bigger, up to thousands of bases, but we used this formula because as explained below the scaling of the entropy loss due to the formation of a loop has the same dependence on the length of the loop as in classical polymer theory. The logaritmic factor on the 2.56 accounts for two phenomena. The first is that a hairpin can bee seen as made by two complementary ssDNA sequences at an effective concentration. This concentration is $c_{eff} = \frac{1}{V} = \frac{1}{d_F^3}$ where $d_F$ is the end to end distance of a chain of the same length of the loop. If the loop (made by L bases) is bigger than the the number n of complementary bases (i.e. $L >> bn$) $d_F$ depends only on L. The melting temperature of a duplex at the concentration $c_{eff}$ is then:

$$Tm = \frac{\Delta H_0}{\Delta S_0 + Rln(c_{eff})} \tag{2.57}$$

when $\Delta H_0$ and $\Delta S_0$ are the enthalpic and entropic costs that depend on the kind of bases involved. If we substitute $c_{eff}$, the logaritmic term of 2.57 becomes

$$Rln(c_{eff}) = Rln(\frac{1}{d_F^3}) = -3Rln(d_F) \tag{2.58}$$

For an ideal chain described by a random walk it becomes

$$Rln(c_{eff}) = -3Rln(b) - \frac{3}{2}Rln(L) \tag{2.59}$$

where b is the fixed dimension of the monomers. For a real chain there is a small difference, but the scaling with L is the same:

$$Rln(c_{eff}) = -3\nu Rln(L) \tag{2.60}$$

where $\nu$ is approximately 0,6 for a swollen chain.

The logaritmic term of 2.56 also depends on the entropy loss of the chain due to the formation of a loop. After the hairpin closure the loop monomers have a smaller conformational space than before, so their entropy decreases. In the classical polymer theory the loss of entropy due to the loop formation is described by 2.61

$$\Delta S = \frac{3}{2}ln(L) \tag{2.61}$$

where L is the length of the loop in monomer units [20][21].

The universality of 2.56 allows us to use it for loops of the order of up to thousands of bases, even though it has been experimentally validated only for short loops.

We calculated the $T_m$ of all the possible contacts considered so far for M13mp18 and in Fig.2.21 we plotted the melting temperature as a function of the length of the loop. We considered a salt concentration of 50 mM of NaCl and a starting loop of 3 thymines for the first term of the denominator in 2.56. For the 12 and 11 basepairs contacts we used the exact enthalpic and entropic values of the sequences from [22], as listed in table 2.2. For the other possible contacts we could not use the exact values for the enthalpy and

| sequence | contact length [monomers] | $\Delta H$ [kcal/mol] | $\Delta S$ [kcal/mol] |
|---|---|---|---|
| GTTTCCATTAAA | 12 | -84,6 | -0,253 |
| TCAGACGATTG | 11 | -81,1 | -0.236 |
| AGACGCTCGTT | 11 | -83,1 | -0.235 |
| GGTGCCGGAAA | 11 | -82,9 | -0,232 |
| GGTTCCGAAAT | 11 | -79,9 | -0,231 |
| TGAAATTGTTA | 11 | -75 | -0,226 |
| ATGAATTTTCT | 11 | -74,1 | -0,223 |
| CCTGTTTAGCT | 11 | -78,7 | -0,228 |
| GTTTATTTTGT | 11 | -76,8 | -0,233 |
| TAATTAATTTT | 11 | -71,6 | -0.223 |

**Table 2.2:** Enthalpy and entropy variation [22] for the closure of the intra-chain contacts of 12 and 11 couples of bases and loop of three non interacting bases of M13mp18. Only one half of the two complementary single stranded sequences is shown as contact sequence. The ionic strength at which the parameters have been calculated is 50 mM NaCl.

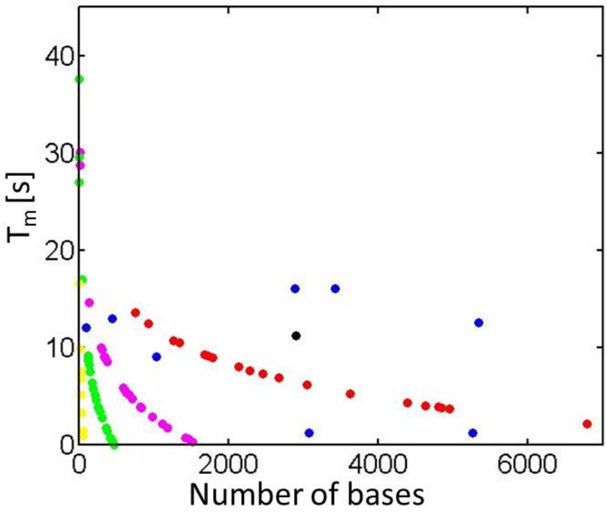


**Figure 2.21:** Melting temperature of the possible contacts of M13mp18 as a function of the length of they loop they would form. Contacts that involve a different number of bases are plotted in different colors. The color code is the same of 2.3

entropy change because there were too many of them, so for every contact length we considered the $\Delta H_x$ and $\Delta S_x$ corresponding to an average duplex made by 50% of GC couples equally distributed along the sequence.

If we look at Fig.2.21 we notice that at 70°C no contact is closed and the ssDNA is a swollen coil because this temperature is higher than the melting temperature of any contact. If we decrease the temperature down to 60°C we reach the melting temperature of a 11 base pair sequence between two regions distant 100 bases along the chain. This contacts closes but does not affect much the conformation of the chain because the regions involved are very close. If the temperature is decreased till 20°C there are other contacts that close but all of them involve close regions of the chain. At about 16°C there are two contacts that involve 11 basepairs at a distance of about 3000 and 3500 bases along the chain (see table 2.3).

| contact name | base sequence | distance [monomers] | Tm [°C] |
|---|---|---|---|
| $\alpha$ | TCAGACGATTG | 3438 | 16.03 |
| $\beta$ | AGACGCTCGTT | 2898 | 16.01 |

**Table 2.3:** The two contacts with the highest melting temperature that involve regions of the chains that are more than 2000 bases far away along the chain. The reported base sequence is the sequence of one of the two complementary regions forming the contact. The other is complementary. The distance is the number of monomers (nucleotides) that form the loop between the two contact regions. $T_m$ is calculated as in 2.56.

The closure of these contact strongly modifies the conformation of the chain and brings closer points of the chain that were far apart before. The structure of the chain at 15°C should resemble Fig.2.22. We called A the first nucleotide of the ssDNA, B the 285th nucleotide (one region of contact $\alpha$), C the 2004th nucleotide, that correspond to half of the loop formed by contact $\alpha$, D the 3723th nucleotide (the second region of contact $\alpha$), E the 3920th nucleotide (first region of contact $\beta$), F the 5369th nucleotide that indicates half of the loop formed by contact B and G the 6818th nucleotide of the chain that is the second region involved in contact $\beta$ (see Fig.2.22).

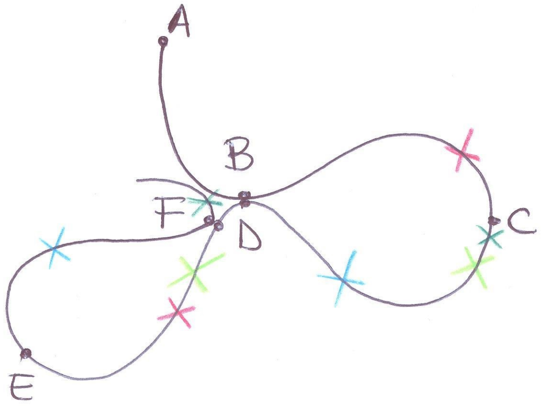


**Figure 2.22:** Sketch of the conformation of M13mp18 at 15°C after the closure of the two contacts ($\alpha$ and $\beta$, table 2.3) involving 11 base pairs. The most distant regions of the chain that close while decreasing the temperature from 15°C to 10°C are shown by crosses. Firstly the magenta contact closes, then the dark green, the light blue and then the light green.

It is not easy to calculate the theoretical dimension of a chain in such conformation, but we can approximate the chain as a Gaussian star polymer with three arms made by 2400 monomers that is the average length between the half lengths of the two loops and the linear piece sequence from base 1 to point B of Fig.2.22. From the classical theory of Gaussian chains [14] the hydrodynamic radii $R_H$ of linear and star polymers can be calculated as in 2.62.

$$R_{H,lin} = \left[\frac{128}{3\pi}\frac{1}{b\sqrt{N}}\right]^{-1} \quad R_{H,star} = \left[\frac{128}{3\pi}\frac{1}{b\sqrt{N}}\frac{(\sqrt{2}-1)(\sqrt{2}+n_A)}{n_A}\right]^{-1} \tag{2.62}$$

where in the linear polymer N is the number of monomers while in the star N stands for the number of monomers of each arm of the star polymer, $n_A$ is the number of arms of the star polymer and b is the Kuhn length, that is 1.5 A for ssDNA. we calculated the $R_H$ of a linear polymer made by 7429 bases that corresponds to a polymer chain of 2899 monomers with a Kuhn length of 1,5 A and we obtained 22 nm. The $R_H$ of a star polymer of 3 branches of 2400 monomers each (i.e. 960 monomers with a Kuhn length of 1,5 nm) is 20 nm. The ratio between the two is 94%, that has to be compared with the measured ratio between the $R_H$ at 15°C and 30°C that is 77%, while between 10°C and 30°C is 70%. The two ratios are not compatible, but this difference could be due to the rough approximation of the two loop as two single stranded sequences of half of the length.

The closure of contacts $\alpha$ and $\beta$ causes the compaction of the chain, and as a consequence changes the melting temperature of many contacts. In this configuration we recalculated the melting temperature of still open contacts and we noticed (Fig.2.23) that some of them increase their $T_m$ because the loop length decreases . In fact it becomes the smallest distance along the chain between the two regions of a contact. For example a contact between 10 base pairs between the sequences 3329-3339 and 273-283 has a $T_m$ of 6°C before the closure of contacts $\alpha$ and $\beta$ because the loop that should form consisted of 3056 bases. After the closure of $\alpha$ and $\beta$ the loop that the 10 bp contact should form is decreased to $l = (D - 3329) + (273 - B) = (3723 - 3329) + (283 - 285) = 396$ bases if we neglect the length of the pairing region. The melting temperature increases and becomes 23°C.

The loop calculation was implemented as follows. For every possible contact we listed the two regions involved by indicating only the nucleotide at the 5′ end of each region. We called $r_1$ the region whose nucleotide number was bigger and $r_2$ the other. Then we considered different cases as $r_1$ and $r_2$ were located in different positions of the chain.

1. If $r_1 < B$ the loop length expressed in units of number of bases is $l = r_2 - r_1 - n$ where n is the number of bases involved in the contact. Typically n is negligible respect to l, so for sake of simplicity we will write just $l = r_2 - r_1$ from now on.
2. If $B < r_1 < D$ and
   - $B < r_2 < D$ the contact introduces two loops in the chain, whose lengths are $l_1 = r_1 - r_2$ and $l_2 = (D - r_1) + (r_2 - B)$
   - $r_2 < B$
     - if $r_1 - r_2 > (D - B)/2 + (B - r_2)$ then $l = (D - r_1) + (B - r_2)$
     - if $r_1 - r_2 > (D - B)/2 + (B - r_2)$ then $l = r_2 - r_1$
3. $D < r_1 < E$ and
   - $C < r_2 < E$ then $l = r_2 - r_1$
   - $B < r_2 < C$ then $l = (r_1 - D) + (r_2 - B)$

  - $r_2 < B$ then $l = (r_1 - D) + (B - r_2)$

4. $E < r_1 < F$ and
    - $C < r_2 < F$ then $l = (r_1 - D) + (r_2 - B)$
    - $B < r_2 < C$ then $l = (r_1 - D) + (r_2 - B)$
    - $r_2 < B$ then $l = (r_1 - D) + (B - r_2)$
5. $F < r_1 < G$ and
    - $F < r_2 < G$ then $l = r_2 - r_1$
    - $E < r_2 < F$ then two new loops form, whose lengths are $l_1 = r_1 - r_2$ and $l_2 = (G - r_1) + (r_2 - E)$
    - $C < r_2 < E$ then $l = (G - r_1) + (E - r_2)$
    - $B < r_2 < C$ then $l = (G - r_1) + (E - D) + (r_2 - B)$
    - $r_2 < B$ then $l = (G - r_1) + (E - D) + (B - r_2)$
6. $G < r_1 < H$ and
    - $F < r_2 < G$ then $l = r_2 - r_1$
    - $E < r_2 < F$ then $l = (r_1 - G) + (r_2 - E)$
    - $C < r_2 < E$ then $l = (r_1 - g) + (E - r_2)$
    - $B < r_2 < C$ then $l = (r_1 - G) + (E - D)(r_2 - B)$
    - $r_2 < B$ then $l = (r_1 - G) + (E - D) + (B - r_2)$

All the contacts whose melting temperature becomes higher than contacts $\alpha$ and $\beta$ after the closure of $\alpha$ and $\beta$, immediately close after them. These contacts should close just after the closure of contacts $\alpha$ and $\beta$ at 16°C and should increase the compaction of the chain. It is not easy to predict which of these contact effectively closes because every closure implies the formation of a segment of double helix and some very close regions of the chain are involved in more than one contact.

## 2.11 Conclusions

We have investigated the coil to globule transition of a single-stranded chain of DNA taken from a viral plasmid (M13mp18) as a model of heteropolymer. Firstly we characterized the distribution of the interactions (contacts) within regions of the same polymer comparing it to a random sequence. Specifically, in section 2.1 we studied the number of possible contacts as a function of the number of nucleobases involved, and the distribution of the interacting sequences along the chain. We found no peculiar structures that could distinguish M13mp18 from a random sequence with the same percentage of A, T, G and C bases.

Secondly we measured with Dynamic Light Scattering and Fluorescence Correlation Spectroscopy the hydrodynamic radius of M13mp18 at different temperatures and with different procedures. In sections 2.9.1 and 2.9.2 the results are shown. We found two characteristic dimensions, around 20 nm at low temperatures and approximately 30 nm at high temperatures, that are connected by a transition that occurs at about 15°C. The amplitude of the temperature transition region is found to depend of the rate of temperature decrease/increase. This difference brings to the conclusion that the folding and

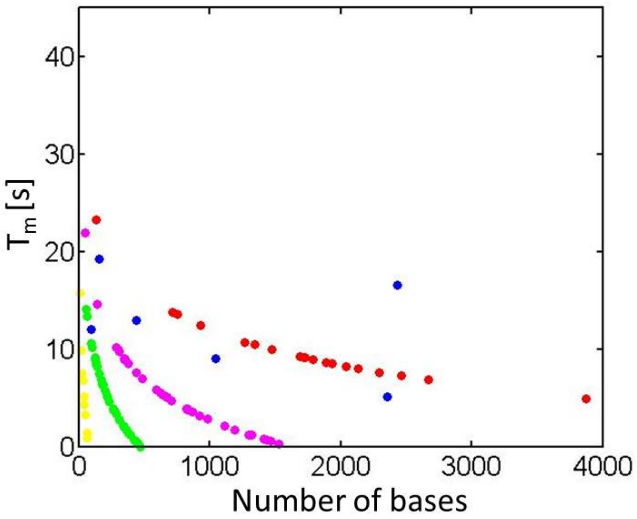


**Figure 2.23:** Melting temperature of the possible contacts as a function of the length of the loop after the closure of contacts $\alpha$ and $\beta$. The color code is the same of 2.3

unfolding kinetics depend on the rate of temperature decrease/ increase, but we could not explore this issue so far.

Thirdly, we tried to interpret the thermodynamics of the observed coil to globule transition of M13mp18 in order to understand which contacts close at a given temperature within the experimental range (see section 2.10). We listed all the possible pairs of segments of the chain that could interact together to form segments of double helix, and we calculated for each of them the melting temperature. This quantity is the temperature below which the corresponding contact closes. It depends on the length on the segments involved and on the distance $l$ along the chain of the two complementary segments. We built a map of the melting temperature as a function of $l$ for all the possible contacts and we found that at about 15°C there are two contacts that involve a large number of bases and have the highest melting temperature though having a large $l$, of about 3000 bases. The closure of these two contacts causes a big compaction of the polymer because they involve segments of the chain that are far apart along the chain. This discovery is compatible with the experimental observations having a transition at the same temperature.

CHAPTER 3

# Liquid crystals

The liquid crystalline LC phase is an intermediate phase between the crystalline and isotropic phase. The molecules of a crystal are characterized by a long range 3-D order, while the molecules of an isotropic liquid are not arranged in any ordered way, nor even locally. Instead the molecules composing a liquid crystal are disposed in a partial ordered way. Typically a molecule that forms LC phases can be thought to have some shape anisotropy, such as an elongated or discoidal shape. These kind of molecules are often composed by a rigid part that is responsible for the steric interactions that brings to the partial ordering, and also by a flexible part that confers fluidity to the phase. Moreover also solutions of amphiphilic molecules form LC crystalline phases. The reason of the formation of such ordered phases has to be found in the structure of the molecules, that have a polar head and an hydrophobic tail. These molecules interact one with the other forming linear aggregates that have the hydrophobic components on the internal part and the polar ones at the outer part. Such discoidal aggregates then interact together to minimize the exposition of the hydrophobic parts to water and form long aggregates that exhibit LC ordering at some conditions.

## 3.1 Some liquid crystalline phases

The liquid crystalline phases can be organized in four big categories, often visible in the same substance with different environmental conditions. There is a commonly used classification of the LC phases based on the number of translational correlations between the molecules. The phase without translational correlations is called nematic phase. If the correlations are one-dimensional than we have the smectic phase; when the correlations become two-dimensional we find the columnar phase and then there are other LC phases with 3-D correlations.

### 3.1.1 Nematic phase

The most simple kind of nematic phase is the uniaxial nematic, shown in Fig.3.1.

In a uniaxial nematic phase the molecules have an average orientation around the nematic director **n**. It is not a crystalline phase, so the orientation is not the same for all the molecules but there is an average direction around which they fluctuate. The degree of order of a nematic phase is quantified by the order parameter $S$, that is defined as follows:

$$S = \frac{1}{2}\langle 3\cos^2\vartheta - 1\rangle \tag{3.1}$$

where $\vartheta$ is the angle between the nematic director and the longitudinal axis of each molecule. The average in 3.1 is a statistical average, calculated over all molecules and

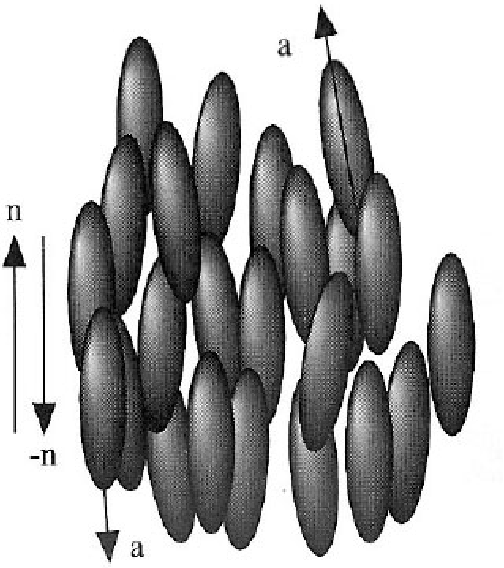


**Figure 3.1:** Uniaxial nematic phase where **a** is the axis of one molecule and **n** is the nematic director.

time. $S$ of a isotropic fluid would be none, while it would become 1 in a perfectly ordered phase. For a nematic phase $S$ assumes values in between $0$ and $1$.

If the molecule composing the nematic phase are chiral, than also the nematic phase becomes chiral and is called cholesteric ($N^*$). The only difference between cholesteric and nematic phase is the presence of a macroscopic chiral structure that originates spontaneously due to the same chirality of the molecules. For example if the molecules are helical then the nematic phase they form develops a helical structure whose principal axis is perpendicular to the nematic director. This cholesteric phase can be thought as formed by nematic layers which are tilted one from the other like shown in Fig.3.2. The direction of the nematic vector is different for every layer and can be described by some simple relations, if we consider the axis $\chi$ as in Fig.3.2:

$$\mathbf{n} = \mathbf{n}(\chi) = \begin{pmatrix} \cos(2\pi\chi/P + \phi_0) \\ \sin(2\pi\chi/P + \phi_0) \\ 0 \end{pmatrix} \tag{3.2}$$

where $P$ is the pitch of the macroscopic helix. it is positive for a right-handed helix and negative for a left-handed one. $\phi_0$ is a constant that depends of the conditions at the boundaries.

### 3.1.2 Smectic phase

The smectic phase is characterized by a almost full one-dimensional translational order and by a sinusoidal distribution of the centers of mass of the molecules. As we can see in Fig.3.3, the molecules composing the phase form compact layers in which they are on average parallel to each other, then in the spacing between the layers they behave like a fluid.

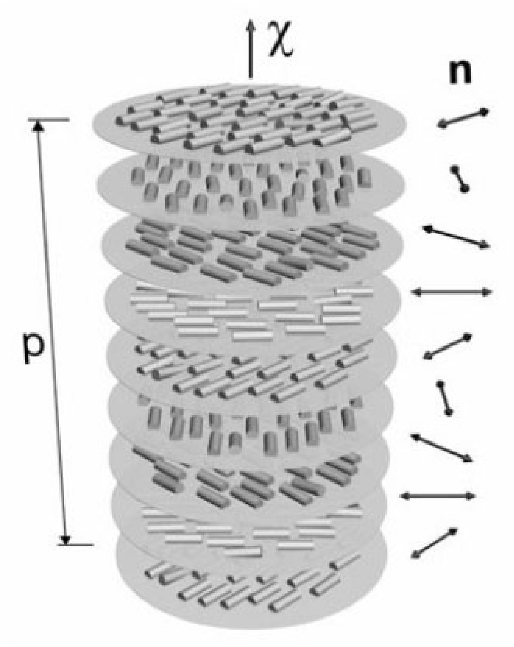


**Figure 3.2:** Cholesteric phase formed by helical molecules. The direction of the nematic vector is shown.

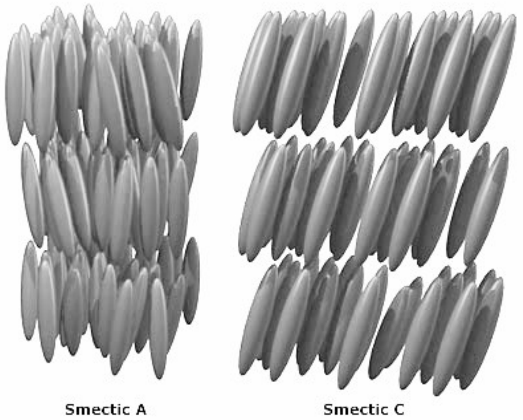


**Figure 3.3:** Smectic phase of kind A and C, characterized by two different average directions of the molecules.

### 3.1.3 Columnar phase

The columnar phase is observed in solutions of lyotropic or discoidal molecules like chromonics. Typically the molecules are disposed one over the other in different columns but with a variable distance between them. The most common columnar structure is the hexagonal one, shown in Fig.3.4, but there other less probable structures. The columnar phase is different from a crystal because the columns that are formed can slid one from the other and every molecule is free to rotate around its main axis.

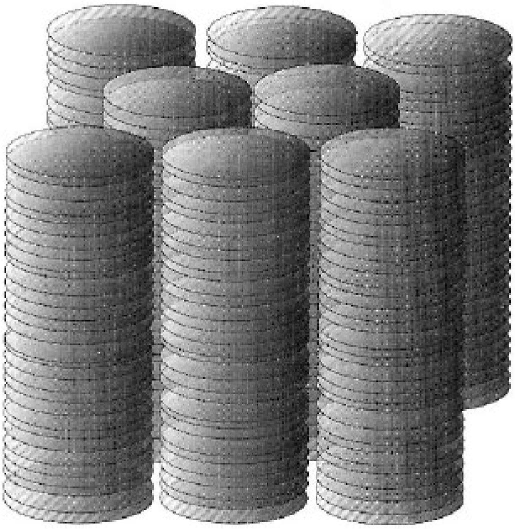


**Figure 3.4:** Example of a hexagonal columnar phase.

## 3.2 Formation of a LC phases

The formation of a liquid crystalline phase in a sample of molecules is caused by the anisotropy of the molecules and by the steric interactions. Usually a sample exhibits a simple transition between the isotropic and the LC phase. This transition has been understood for the first time by Onsager in 1949 studying a suspension of stiff cylindrical rods that interact via steric repulsion only. At fixed temperature the free energy of the system has just an entropic contribution. In the isotropic phase the rods all directions with the same probability so the total entropy of the sample seems to be higher than the entropy of the same sample ordered in a LC phase. However we must consider that, within a certain concentration range, in the LC phase the rods can translate more freely than in the isotropic phase, and the excluded volume for a rod due to the presence of a second one is minimum in this situation. Consequently above a critical concentration the entropic contribution related to the LC phase becomes higher than the isotropic contribution of the isotropic phase, so the LC phase becomes favourite than the isotropic. Onsager demonstrated that the phase transition that the system undergoes when it reaches the critical concentration is of the first order. The critical concentration for a sample made of elongated molecules can be expressed through the critical volume fraction $\phi$, that is the ratio between the total volume of the molecules and the volume of the solution. $D$ is the diameter of the molecules, $L$ their length and $c$ is the concentration that is calculated as the ratio between the number of molecules and the total volume of the solution.

$$\phi = c\frac{1}{4}\pi L D^2 \tag{3.3}$$

Onsager theory applies to a simplified situation in which the molecules interact only via steric repulsion. In a real sample there are always many complicated interactions, but some of them in principle can be modelled with the steric interaction. Some necessary conditions for the formation of the LC phases are that the volume fraction $\phi$ of the sample must be small ($\phi \ll 1$) and that the elongation of the molecules must satisfy the following relation: $L \gg D$. From these two conditions we can calculate the smallest volume fraction that is necessary to observe a nematic phase in a sample of elongated molecules of diameter $D$ and length $L$: $\phi \sim 4D/L$.

There are several computational models that apply Onsager theory to different kind of molecules. Starting from the steric interactions, they progressively introduce different and complicated interactions between the molecules or even peculiarities of the shape of the molecules, such as polydispersity of the sample or flexibility. Studies were made on solutions of charged elongated molecules, because the electrostatic interaction can be modeled with the steric interaction just in the isotropic phase, while when the molecules have some average orientation they interact differently depending on their relative orientation. Thus the electrostatic interactions increase the critical concentration $c_{I-N}$ above which the nematic phase if formed and reduces the order parameter of the nematic phase in isotropic-nematic coexistence. If we introduce some flexibility of the molecules we notice that $c_{I-N}$ increases, and also the concentrations of the smectic and columnar phases. The polydispersity suppress the formation of the smectic phase because it is difficult to obtain a one-dimensional translational order with molecules with different lengths.

### 3.2.1 Living polymers

Many kind of molecules like some proteins or chromonics, that don't have sufficient shape anisotropy for the formation of LC phases, when put in a solution spontaneously form elongated aggregates that become long enough to have the anisotropy required by Onsager rules. It has been observed that for some molecules at low temperatures the free energy associate to the polymerization process favors the aggregation of the monomers. This peculiar kind of molecules are called living polymers. They usually form LC phases because the formation of elongated aggregates yield the ordering of the phase, but also the nucleation of an even partially ordered phase helps the polymerization process. This interplay between polymerization and order is fundamental in understanding the formation and the stability of the LC phases of DNA.

It is possible to define a melting temperature of the nematic phase of a solution of living polymers, that is the temperature at which the sample undergoes from a nematic to an isotropic phase. The melting temperature of the phase provides a tool to measure the binding energy between the molecules and gives an estimate of the length of the aggregates. It is possible to find an analytic expression for the average number of molecules per aggregate $< i >$ (see equation 3.4) that depend on the density $\rho$ of the sample, i.e. the ratio between the number of molecules and the total volume of the solution, $v_0$ that is the accessible volume for every molecule if the consecutive is kept fixed, $\Delta S$ is the entropy variation due to the polymerization of the system, $\beta = 1/k_B T$ and $\varepsilon$ is the absolute value

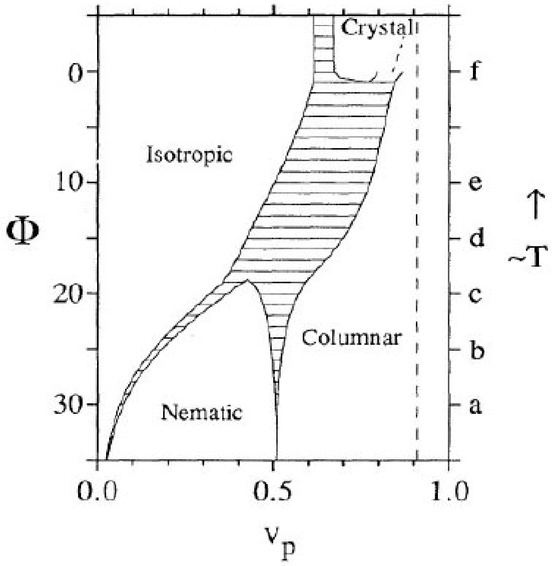


**Figure 3.5:** Phase diagram of self aggregating elongated molecules. $\phi$ is the volume fraction of the molecules, $v_p$ the volume of the single molecule and $T$ the temperature of the system. Shadowed areas indicate the coexistence region, while the vertical dotted lines are the steric limit for the packing. Points from $a$ to $f$ stand for different effective temperatures of the sample.[23]

of the binding energy between a couple of molecules.

$$< i >= \frac{1}{2}\left(1 + \sqrt{1 + 4\rho v_0 e^{-\Delta S/k_B} e^{\beta\varepsilon}}\right) \tag{3.4}$$

## 3.3 LC phases of DNA

From the '40s it has been known that the double helix of DNA in aqueous solutions forms LC phases. This fact plays also a fundamental role in the discovery of the shape of double stranded DNA. In fact, the columnar ordering of long helices of DNA where every column of duplexes is free to translate along its axis, enhances the molecular structure factor when observed through X-rays scattering. Many studies were done on solutions of long (from N=$10^7$ to N=100 base pairs) double stranded DNA molecules, that revealed the formation of isotropic ($I$), chiral nematic ($N^*$), columnar ($C$) and crystalline ($X$) phases at increasing concentration. The formation of such phases can be explained through Onsager theory: he described a system of elongated monodispersed molecules of length $L$ and diameter $D$ that develop LC phases. If they are sufficiently elongated they can order to form a nematic phase if their volume fraction $\phi > \phi_{IN} = 4D/L \sim 247N$, where $D \simeq 2nm$ and $L \simeq N/3nm$ for a B-form of DNA. Bolhuis and Frenkel performed some simulations on a system of hard rods and confirmed this prediction from a quantitative point of view for $L/D > 4,7$ (i.e. a DNA double helix of 28 base pairs), and they also show that for $L/D < 4,7$ there are no LC phases for any $\phi$ value. If we superimpose the experimentally measured phase diagrams of a solution of long DNA (lDNA, made by approximately 150 base pairs) to the simulated phase diagram we can observe an almost quantitative compatibility with Onsager theory.

When observed by an optical microscope the cholesteric phase of lDNA exhibit a variety of different textures:

- when we observe a portion of phase whose cholesteric axis is parallel to the plane of the sample, we see some periodic fringes called fingerprints. The periodicity is due to the continuous rotation of the axis of each molecule around the sample plane, like shown in Fig.3.6 and 3.7 on the left. The period of the modulations of intensity corresponds to half of the cholesteric pitch, as better explained in chapter 2.6.1.
- when the cholesteric axis is perpendicular to the plane of the sample we observe textures that can be characterized by precise color if the cholesteric pitch is in between $0,2$ and $0,4\mu m$ (see Fig.3.2 on the left).

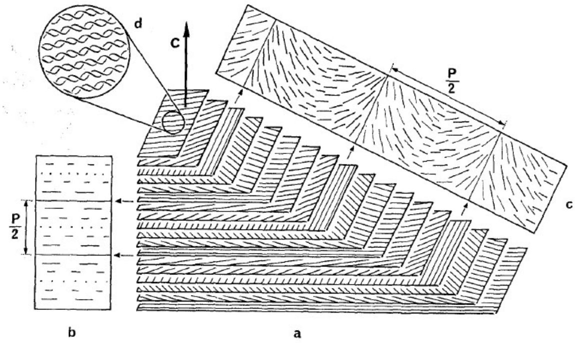


**Figure 3.6:** (a) Scheme of the cholesteric phase of a solution of lDNA. (b) Projection of the nematic director along the direction perpendicular to the helix axis, (c) section along the cleaving plane and (d) representation of the duplexes of the superior layer. [24]

## 3.4 LC phases of short duplexes of DNA

It has been recently discovered that also short DNA duplexes (from 6 to 20 base pairs) show LC ordering [25] [26]. For these molecules the ratio between the length and the diameter is much smaller than the smallest ratio necessary for the formation of LC phases predicted from Onsager theory. The formation of LC phases in a solution of not enough elongated molecules cannot be explained just on the basis of Onsager theory, but it is necessary to assume that the molecules interact and form long linear aggregates that behave like molecules with the same diameter than before but with bigger length. Upon changing the composition of the solvent or the temperature, the phase is destroyed and then restored, so the assembly process is reversible. LC phases of many short DNA sequences were studied (see Fig.3.8) and in particular the nematic phase was well characterized. The sequences studied can be divided into two categories as a function of the conformation of the terminals. In fact many sequences of Fig.3.8 are self-complementary, as shown for example in Fig.3.9: such sequences are named blunt ended sequences (BE) in opposition to the sticky ended sequences (SE). The difference between the two kind of

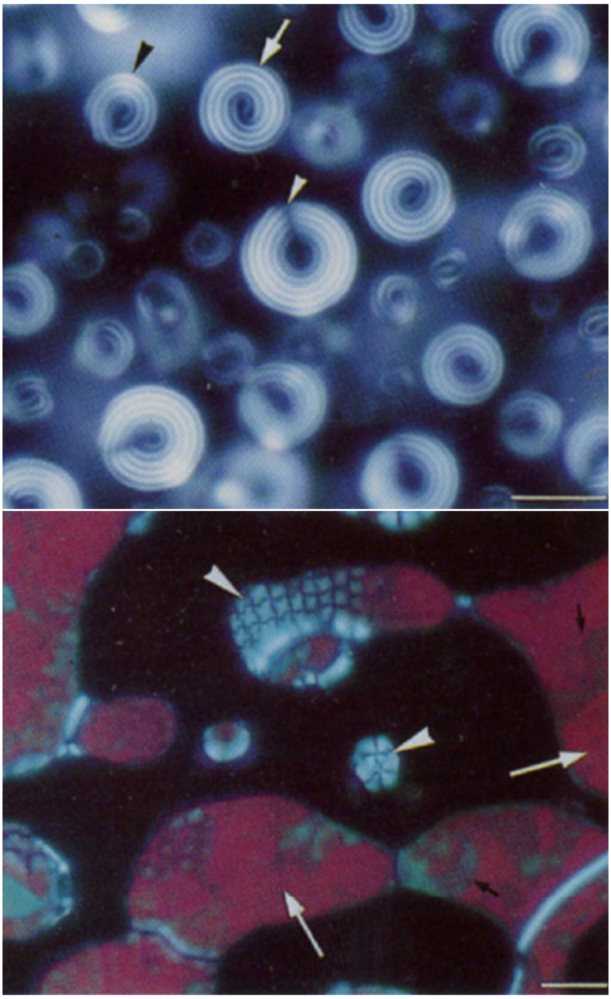
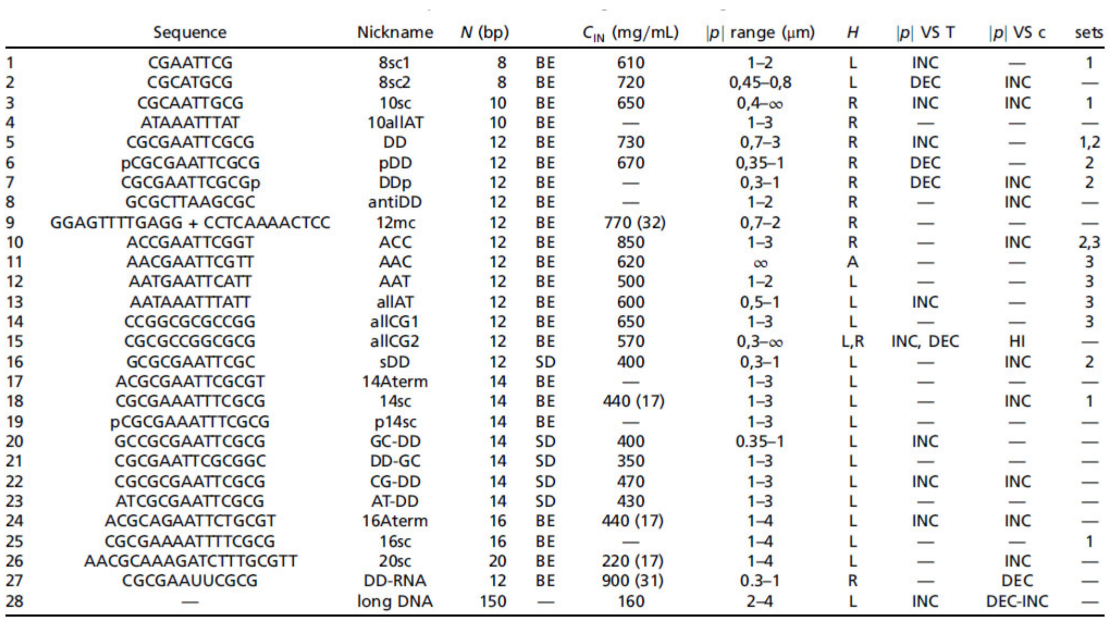


**Figure 3.7:** Example of cholesteric texture of a solution of lDNA when the cholesteric axis is parallel to the plane of the sample (on the left) or perpendicular (on the right). [24]

sequences emerges clearly in the process of polymerization between two duplexes. For SE duplexes the end of one molecule can bind to the end of another molecule and the energy variation due to the closure of the bond consists of two terms, one comes from the pairing interaction and one from the stacking interaction. Instead for BE sequences the bond between two duplexes is weaker because there is no pairing contribution.



| | Sequence | Nickname | *N* (bp) | | $C_{IN}$ (mg/mL) | $\|p\|$ range (μm) | *H* | $\|p\|$ VS T | $\|p\|$ VS c | sets |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | CGAATTCG | 8sc1 | 8 | BE | 610 | 1–2 | L | INC | — | 1 |
| 2 | CGCATGCG | 8sc2 | 8 | BE | 720 | 0,45–0,8 | L | DEC | INC | — |
| 3 | CGCAATTGCG | 10sc | 10 | BE | 650 | 0,4–∞ | R | INC | INC | 1 |
| 4 | ATAAATTTAT | 10allAT | 10 | BE | — | 1–3 | R | — | — | — |
| 5 | CGCGAATTCGCG | DD | 12 | BE | 730 | 0,7–3 | R | INC | — | 1,2 |
| 6 | pCGCGAATTCGCG | pDD | 12 | BE | 670 | 0,35–1 | R | DEC | — | 2 |
| 7 | CGCGAATTCGCGp | DDp | 12 | BE | — | 0,3–1 | R | DEC | INC | 2 |
| 8 | GCGCTTAAGCGC | antiDD | 12 | BE | — | 1–2 | R | — | INC | — |
| 9 | GGAGTTTTGAGG + CCTCAAAACTCC | 12mc | 12 | BE | 770 (32) | 0,7–2 | R | — | — | — |
| 10 | ACCGAATTCGGT | ACC | 12 | BE | 850 | 1–3 | R | — | INC | 2,3 |
| 11 | AACGAATTCGTT | AAC | 12 | BE | 620 | ∞ | A | — | — | 3 |
| 12 | AATGAATTCATT | AAT | 12 | BE | 500 | 1–2 | L | — | — | 3 |
| 13 | AATAAATTTATT | allAT | 12 | BE | 600 | 0,5–1 | L | INC | — | 3 |
| 14 | CCGGCGCGCCGG | allCG1 | 12 | BE | 650 | 1–3 | L | — | — | 3 |
| 15 | CGCGCCGGCGCG | allCG2 | 12 | BE | 570 | 0,3–∞ | L,R | INC, DEC | HI | — |
| 16 | GCGCGAATTCGC | sDD | 12 | SD | 400 | 0,3–1 | L | — | INC | 2 |
| 17 | ACGCGAATTCGCGT | 14Aterm | 14 | BE | — | 1–3 | L | — | — | — |
| 18 | CGCGAAATTTCGCG | 14sc | 14 | BE | 440 (17) | 1–3 | L | — | INC | 1 |
| 19 | pCGCGAAATTTCGCG | p14sc | 14 | BE | — | 1–3 | L | — | — | — |
| 20 | GCCGCGAATTCGCG | GC-DD | 14 | SD | 400 | 0.35–1 | L | INC | — | — |
| 21 | CGCGAATTCGCGGC | DD-GC | 14 | SD | 350 | 1–3 | L | — | — | — |
| 22 | CGCGCGAATTCGCG | CG-DD | 14 | SD | 470 | 1–3 | L | INC | INC | — |
| 23 | ATCGCGAATTCGCG | AT-DD | 14 | SD | 430 | 1–3 | L | — | — | — |
| 24 | ACGCAGAATTCTGCGT | 16Aterm | 16 | BE | 440 (17) | 1–4 | L | INC | INC | — |
| 25 | CGCGAAAATTTTCGCG | 16sc | 16 | BE | — | 1–4 | L | — | — | 1 |
| 26 | AACGCAAAGATCTTTGCGTT | 20sc | 20 | BE | 220 (17) | 1–4 | L | — | INC | — |
| 27 | CGCGAAUUCGCG | DD-RNA | 12 | BE | 900 (31) | 0.3–1 | R | — | DEC | — |
| 28 | — | long DNA | 150 | — | 160 | 2–4 | L | INC | DEC-INC | — |

**Figure 3.8:** Summary of the fundamental properties of the cholesteric phase observed in solutions of different short duplexes of DNA. For each are listed the name of the sequence, the number of bases, the structure of the terminals, the concentration of the nematic phase, the cholesteric pitch, its behaviour with temperature and concentration and the chirality of the phase. INC stands for increasing, while DEC for decreasing. L determines a left handed phase, and R a right handed phase.[26]

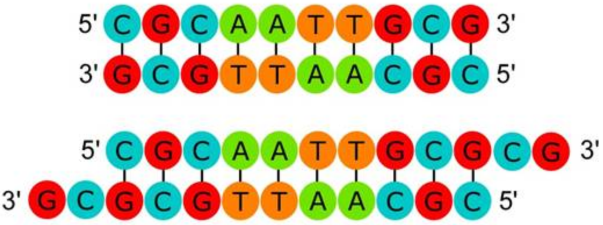

5' C G C A A T T G C G 3'
3' G C G T T A A C G C 5'
5' C G C A A T T G C G C G 3'
3' G C G C G T T A A C G C 5'

**Figure 3.9:** Example of one BE duplex (on the top) and one SE duplex (on the bottom)

The shape of the end of the duplexes influence also the average length of the linear aggregates that form due to the polymerization process. In Fig.3.10 such length is shown ($M$), while $\phi$ is the volume fraction, as a function of SE or BE terminal. $X_0$ is the ratio between the length and the diameter of the duplex, that for example is approximately 2 for 12mers. If we change the conformation of the end of the duplex we change automatically the interaction energy between two duplexes. For BE sequences its value is lower than for SE sequences. From Fig.3.10 we can notice that the aggregates are composed by a small number of duplexes till $\phi \leq 0,3 - 0,4$, then as $\phi$ is increased the aggregates increase suddenly to form the nematic phase.

The nematic phase that is observed in the different DNA sequences listed in Fig.3.8can be classified on the basis of the chirality and the length of pitch at a fixed temperature. Usually a cholesteric phase is left handed if going through it towards the observer in the direction perpendicular to the nematic director, i.e. in the same direction of the helical axis, the nematic director is seen to rotate clockwise. On the contrary, a circularly polarized beam of light is left handed if the electric field that goes through the observer rotates counter-clockwise.

Fig.3.8 shows that cholesteric phases of short DNA duplexes have a wider variety of pitches and handedness than the phase formed by long DNA. The reason of such variety is still unknown because it is not straightforward to predict the characteristics of the cholesteric phase from the sequence of DNA from a known phase. It is not still clear how to predict the phase from the sequence of DNA for the already studied sequences of Fig.3.8. There are some models that try to connect sequence and phase; the classical model developed by Straley considers only the steric interactions (see Fig.3.11) between the molecules and predicts the cholesteric phase to have the opposite handedness than the observed one. Recently some models were developed that consider also the electrostatic interactions between the molecules [27]. Such models allow to calculate the energy of the interaction between two helices as a function of the reciprocal orientation following two strategies: the duplexes are considered as cylinders with a helicoidal chains of electrostatic charges and counterions along the side, or they are modelled as charged helices. The handedness of the cholesteric phase of a solution of lDNA is correctly predicted only by the second model. In these models the electrostatic and steric contributions are represented by two constants, $k_t$ that represents the chirality of the phase and $k_{22}$ that is elastic and tries to get the helices parallel. The ratio between the two constants is connected to the cholesteric pitch:

$$p = -2\pi \frac{k_{22}}{k_t} \tag{3.5}$$

In the second model in particular the competition between the two forces is enhanced. The chiral forces yield to a decrease of the cholesteric pitch, while the elastic forces restore the parallelism between the helices. The elastic term does not show a strong dependance from the peculiar configuration of the microscopic structure of the duplexes. For the chiral forces the dependances are the opposite. In $k_{22}$ the elastic forces are represented, while in $k_t$ both elastic and steric contributions are present. In Fig3.12 are represented both contributions that yield to the determination of the value of $k_t$. The two contribution have an opposite effect on the constant.

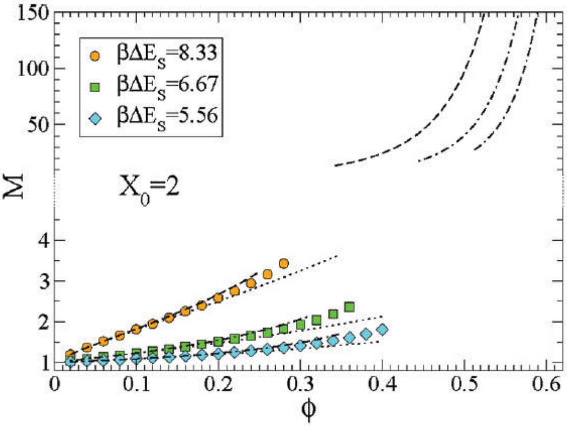


**Figure 3.10:** Average length of the aggregates in monomer units as a function of $\phi$ for sequences of about 12 couples of bases and for different values of the interaction energy between the duplex ends. Symbols come from MC simulations, while the dotted lines from theoretical models. [28]

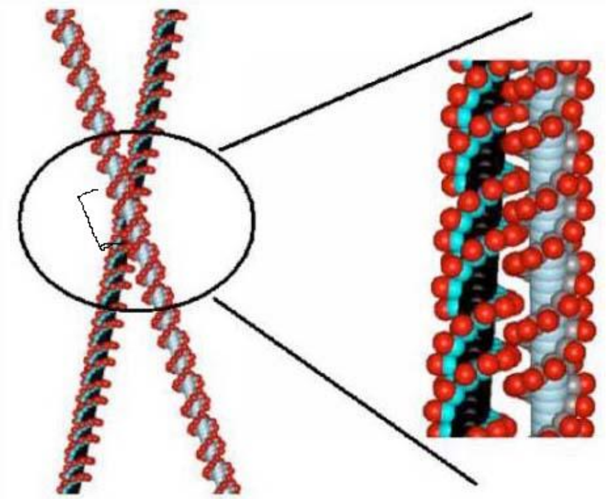


**Figure 3.11:** Steric interactions between two duplexes. [27]

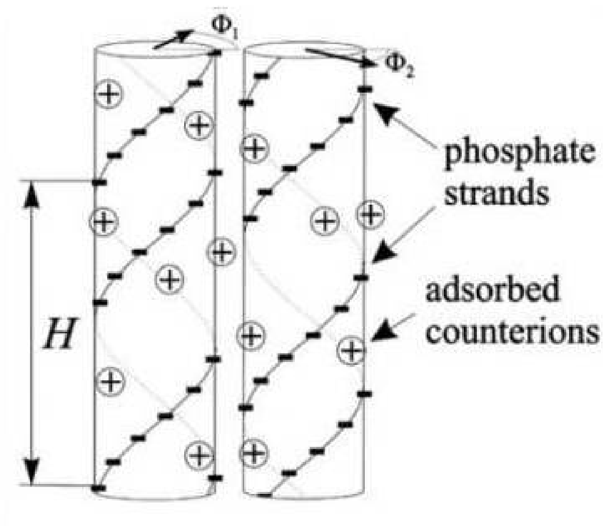

$\Phi_1$
$\Phi_2$
phosphate strands
adsorbed counterions
$H$

**Figure 3.12:** Scheme of a duplex of DNA as a cylinder with helical lines of negative electric charges and positive counterions on the side.

CHAPTER 4

# Liquid crystal ordering of B-DNA duplexes: extracting the interactions from the phase boundaries.

The discovery of LC ordering of DNA oligomer duplexes introduced and described in chapter 3 of this thesis brought about the notion that stacking interactions between distinct blunt ended duplexes is strong enough to guide supra molecular ordering. At the time of the discovery this finding appeared somewhat surprising since there are no clean examples of the effects of such stacking in biological observations. At the same time this discovery brought in awareness that, while the stacking interaction between coupled bases within each single duplexes is quite well characterized, the stacking interaction between independent duplexes is currently not well known [29][30]. A few works are instead available on a somewhat intermediate condition, the stacking interactions within duplexes with a nick usually referred to as coaxial stacking. Despite the contradictory results offered by these two works, they nonetheless clarify that the additional degree of freedom provided by the nick yields a stronger stacking interactions with respect to chemically continuous double helices ($\Delta G \sim 1,4kcal/mol$ for intra-helix stacking at 37°C [22], while $\Delta G \sim 1,7kcal/mol$ with a nick [7]).

Following the discovery of short DNA LC, a few models were proposed to account for the onset of ordering in such weakly interacting monomers. Two independent approaches were developed by modelling the DNA oligomer duplexes as simple geometrics solids, i.e. cylinders [31] or superquadrics [28], which are effectively cylinders with smoothed edges. In both models stacking is introduces as an attractive potential between the terminal bases, which induces linear aggregation of the monomers. The interaction is modelled as a square potential [28] or as truncated polynomial that decays as the distance between the ends of the duplexes is increased [31]. Both models turned to be quite successful in clarifying that for sufficient volume fractions even a moderate attraction could indeed give rise to both nematic and columnar phases. Both models indicate that the ordering transition toward the LC is very discontinuous, more than what predicted and observed in assemblies of rigid rod-like particles. Across the transition, the system not only develops orientational ordering but also undergoes a markedly discontinuous enhancement in the mean length of the aggregates. Therefore the significant effect of the weak stacking interaction is understood as carried by a positive feedback between aggregation and entropically driven ordering. Thus the onset of LC ordering can be effectively considered as a form of amplification of the stacking interactions.

By following a different approach, Maffeo et al [29] investigated the basics of the observed LC ordering through an extended all atom simulations of a solution of DNA duplexes. In particular they carefully investigated the stacking interaction between independent duplexes extracting the potential of mean force as a function of the mutual duplex positions and azimuthal angle. Quite remarkably the interaction strength they obtain ($\Delta G \sim 6kcal/mol$ at $37^{\circ}$C) is markedly larger than the one within continuous helices. They also found a strong dependence of the interaction strength on the azimuthal angle.

The aim of the work described in this chapter was to build on this recently developed knowledge to reverse the process, using the collective behaviour of DNA oligomer duplexes to extract new information on the stacking interaction between independent helices. This is done by combining the coarse grained models with information extracted from the atomistic simulations as follows.

First, we experimentally determined the isotropic-nematic phase boundaries for a model sequence of a DNA oligomer. Specifically, we chose a DNA 12mer ('AllAT') which is known by previous crystallographic investigations to yield semi-ideal B-DNA helices. [32] This choice provided the best experimental approximation to the simplified geometries adopted in the coarse grained models. Moreover, we considered both natural DNA oligomers and racemic mixtures of the enantiomers of AllAT to ensure that the observed phase behaviour was dependent only on stacking and not on potentially complex details of side-side DNA repulsion. The experiments enabled determining the DNA concentration at which LC is formed and the coexistence range as a function of temperature. They also enabled determining the temperature dependence of the nematic order parameter up to the phase melting.

Second, we adopted the model by De Michele et al. to extract from the phase boundaries a simplified characterization of the interaction. The advantage of this model over the one of Glaser [31] is that it enables an explicit computation of the phase boundaries once a set of interaction parameters are adopted. In particular, by this comparison we determined the temperature dependence of the interaction strength and of the persistence length of the aggregates. However, the strength of this model is also its weakness: the knowledge of these two parameters is difficult to map into a characterization of the strength of the stacking interaction, as fully described in this chapter.

Third, we extended the model by De Michele et al. by incorporating into it some information which were made available by the atomistic simulations. Indeed, through a collaborative work with Maffeo and Aksimentiev, we determined the detailed shape of the stacking interactions for AllAT resulting from atomistic simulations. On the basis of this potential we computed the parameters necessary for the calculation of the phase diagram as implemented by De Michele et al. to extract the depth of the stacking interaction as a function of temperature.

The outcome of this research program is two sided. On the one hand we demonstrated that the whole approach is effective, enabling us to extract a specific values for the parameters at play. On the other hand it became apparent that the stacking interaction resulting from atomistic simulations is too narrow ranged to justify our experimental observations. It has been known for quite a while that atomistic simulations overestimate stacking interactions, so in principle this could be due to this effect of narrow ranged interaction. However, as discussed later in the chapter, we are at the moment adopting a distinct interpretation of this observation. To perform the atomistic simulations the terminal bases of the duplexes were artificially locked effectively reducing the range for base to base interactions. This procedure was adopted even in the previous publications [29]. A new set of atomistic simulations performed with free terminal bases is currently under way. We expect that this new degree of freedom will significantly expand the interaction range. Whether or not this will be sufficient to account for the experimental observations and thus to lead to a new experimental determination of the stacking interaction between independent duplexes, will be thus discovered in the near future.

## 4.1 Sequence

The studied sequence is 5′-AATAAATTTATT-3′ (allAT), a self-complementary 12mer. To perform this study we selected a sequence that has a 3D structure very close to the B-form of DNA. The crystallographic shape of AllAT was studied and it was found to be closer to an ideal double helix than other sequences, like for example the Dickerson Dodecamer [32] [27]. In solution the conformation of AllAT is still not determined but we rely of the fact that in sufficiently stabilizing buffers the shape of AllAT should be closer to a classical double helix than other sequences. Another reason why we chose to study AllAT is that this is a self-complementary sequence. It means that two identical strands of AllAT form a blunt ended double helix. If the sequence was not self-complementary we would have had to dissolve two complementary sequences to obtain a solution of duplexes and also to balance their stoichiometric ratio to avoid possible excess of one or the other sequence. Using a self-complementary sequences simplifies the sample preparation process. Moreover it was recently observed that AllAT duplexes in solution formed a cholesteric phase and the concentration at which this happens was determined, but no further characterization was done (see Fig.4.2a)) [26].

We studied the interactions between the ends of duplexes of AllAT through the characterization of the nematic phase they form. To do so we compared the experimental measurements to a coarse grained model predicts the concentrations of formation of a nematic phase of rigid cylinders. Such model is developed for molecules that are not chiral and thus predicts the formation of a uniaxial nematic phase. The AllAT duplexes instead are intrinsically chiral molecules so the expected LC phase is a chiral nematic [33] [34] [35]. The model also predicts the value order parameter as a function of the concentration and the temperature, so we tried to compare the measured parameter to the calculated one. The order parameter is a difficult quantity to measure in a cholesteric phase, so we chose to study a non chiral nematic phase formed by mixing common right-handed (D-DNA) with left-handed DNA (L-DNA) sequences in equal quantities to form racemic allAT mixtures. The L-DNA are produced by NOXXON PHARMA A.G. for pharmacological purposes, and they are made with the same nucleobases and phosphate groups of the right handed, but with the sugar that is mirror symmetric to the conventional deoxyribose. As a result the double helices formed by left handed sequences are the mirror symmetric of the right handed ones [36]. Because of the enantioselectivity of Watson-Crick interactions left- and right-handed sequences do not bind together, so mixtures of allAT enantiomers in solution form left- and right-handed double helices. We had already roughly mixed two enantiomers of the same sequence and we observed that they develop a nematic phase with a small chiral twist [37]. We tried then to carefully balance the quantities of the two enantiomers in order to obtain a uniaxial nematic phase. For this purpose the racemic solutions were assembled with a circular dichroism (CD) device. A solution of right-handed duplexes when measured with CD at 20C has an absorbance spectrum like in Fig.4.1(a), while the left-handed spectrum has symmetric shape (see Fig.4.1). We mixed the two enantiomers in equal quantities and we manually corrected the ratio between the two till we measured a CD spectrum that show no feature attributable to an excess of one enantiomer or the other.

## 4.2 Sample preparation

AllAT sequences were synthesized from Noxxon Pharma AG., HPLC purified, liophilized and then resuspended in our laboratory to be dialized separately in a bath of 5mM NaCl. Then they were liophilized again. After this process, to build the LC samples the AllAT

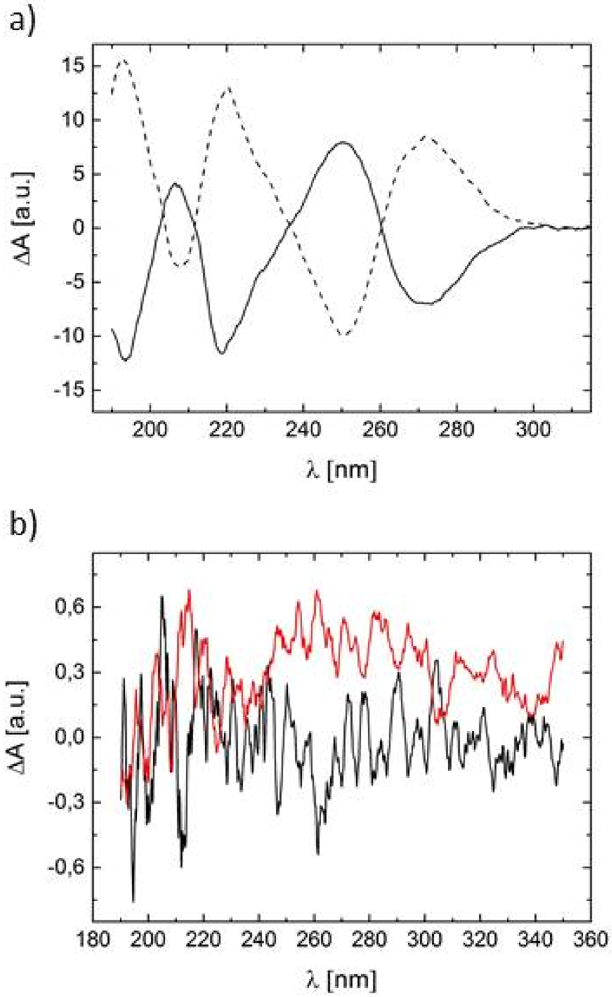


**Figure 4.1:** Circular dichroism spectra of a solution of (a) right-handed AllAT (dotted black line), left-handed AllAT (solid black line), (b) a racemic mixture of the two enantiomers (red line) and the baseline (black line). All the solutions were concentrated at 0,2 g/l DNA.

sequences were dissolved in water at a concentration of about 30 mg/ml with no added salt or with 1,8M of NaCl. They were further concentrated by evaporation, closed in flat cells with d = 20 $\mu$m gap between the glass plates, sealed with fluorinated oil, and aged and thermally cycled to ensure uniformity of concentration within each cell.

## 4.3 AllAT nematic phase

We studied the nematic phases of water solutions made by:

- racemic allAT without NaCl added,
- racemic allAT at 1,8M of NaCl,
- only D-allAT at 1,8M of NaCl,
- only L-allAT at 1,8M of NaCl.

We observed the formation of a chiral nematic phase in both samples of D-AllAT and L-AllAT (see for example Fig.4.2(a)), while the samples of racemic mixtures showed no sign of chirality (see Fig.
refnemallat1(b)).

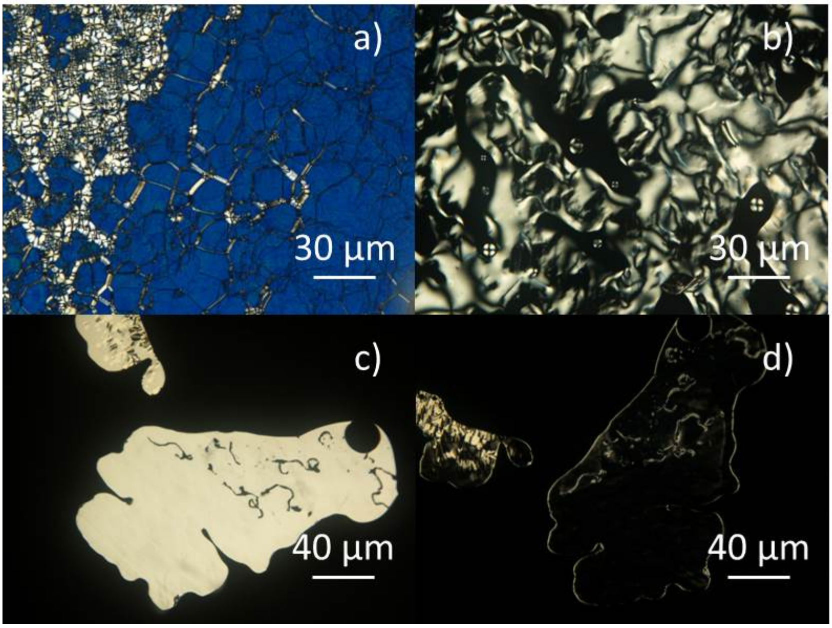


**Figure 4.2:** Polarized optical microscopy image of a nematic phase of a) a solution of D-allAT at a concentration of 595 g/l of DNA with 1,8M of NaCl and b) racemic allAT with 1,8M of NaCl at a concentration of 580 g/l. The images were taken with crossed analy¡er and polarizer, so the black regions are the one with isotropic phase. c) Racemic allAT without added salt at a concentration of 590 g/l d) the same racemic sample rotated by 45°. All images were taken with crossed analyzer and polarizer.

From Figs.4.2b) and 4.2c) we deduce that the phase is nematic and aligned in the plane of the cell. Since intensity of the phase greatly changes by rotating the sample (i.e.

the nematic director) of about 45° between crossed polarizers, almost all the helices are parallel to the analyzer or to the polarizer when the intensity is extinguished, while they are rotated by 45° from the polarizer or the analyzer at the maximum of the intensity.

### 4.3.1 Phase diagram of nematic allAT

For each of the four solutions we prepared many different samples exploring the range of concentrations and temperature where the nematic phase was formed. Within the range of existence of the nematic phase, we measured the melting temperature $T_m$ and the DNA concentration. We defined as the melting temperature the temperature at which the all the phase was melted. It gives an indirect measurement of the strength of the interactions between the duplexes. As the temperature is increased nematic phase of DNA duplexes melts because the stacking interactions between the helices decrease. To measure the $T_m$ we heated (1°C per minute) each sample. For example if a sample was found to be nematic at RT, while increasing the temperature the nematic phase started to melt, till at $t_m$ is became completely isotropic. Indeed there is a temperature range in which we observed the coexistence between the isotropic and the nematic phases. We called $T_0$ the temperature at which the nematic phase started to melt. Then we calculated the coexistence temperature range as $\Delta T = T_m - T_0$.

The DNA concentration was measured for all the samples with an optical microscope and through an Fabry-Perot technique (see chapter 2.6.1).

The nematic phase of racemic mixture of AllAT sequences was studied as first. We observed that at 20°C the phase is isotropic below 500 g/l of DNA while above 550 g/l it is nematic. In between the two concentrations we found a phase coexistence. In Fig.4.3 we plotted the $T_m$ versus the DNA concentration. The straight lines ($T_m(c)$) fits the dataset. Above those lines the phase is isotropic.

The horizontal error bars of Fig.4.3 are the uncertainties on the concentration measurements. They are very large due to the intrinsic difficulty of measuring the concentration. We could not prepare the samples with the exact concentration in order to avoid this measurement because this would require a large amount of sample. Moreover a solution of DNA and water at the concentration necessary to form the nematic phase has a high viscosity and is difficult to handle.

As explained in chapter 2.6.1 the DNA concentration in a nematic phase is here measured from the refractive index of the solution. The origin of this large error on the DNA concentration is in the weak sensitivity of the refractive index on the concentration. Indeed an error of $10^-3$ in the refractive index correspond a large error in the concentration, but unfortunately we could not find a better method to obtain the DNA concentration of the nematic phase. We tried to partially compensate this problem by repeating the measurements a large number of ties in independent prepared cells. The results are plotted in Fig.4.3.

For the temperature range of $\Delta T = T_m - T_0$ (the vertical lines, whose ends are $T_m$ and $T_0$) we observed a coexistence between isotropic and nematic phases, while below $T_0$ the phase is nematic.

We tried to further reduce the noise of the phase diagram by considering the temperature range $\Delta T = T_m - T_0$ of the phase coexistence. When measured the $\Delta T$ is plotted as a vertical line, whose ends are $T_m$ and $T_0$. Between those temperatures we observed a coexistence between isotropic and nematic phases, while below $T_0$ the phase is nematic. In Fig.4.4 we plotted the temperature range of the isotropic-nematic coexistence vs the melting temperature for many different concentrations. The measurements of $\Delta T$ and

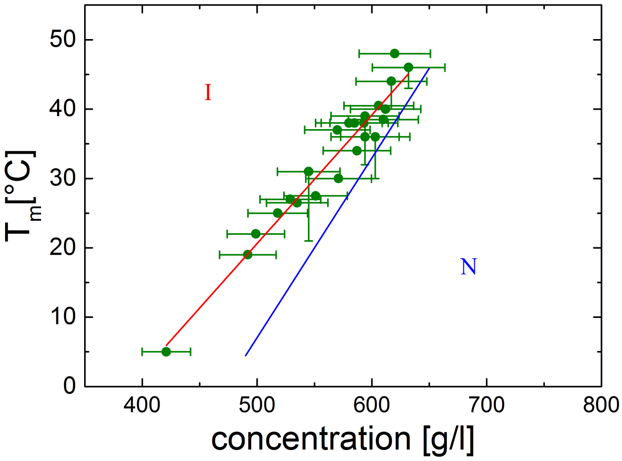


**Figure 4.3:** $T_m$ vs DNA concentration for the racemic mixture of AllAT without added salt. The red straight line is the fitting line of the green dots and provide the phase boundary of the isotropic phase. The blue line is the phase boundary of the nematic phase calculated through the data plotted in Fig.4.4 as explained in the text. In between the blue and the red lines we observed isotropic-nematic phase coexistence.

$T_m$ seem to be affected by smaller errors than the measurement of the DNA concentration. The data were fitted with a straight line, $\Delta T(t_m)$.

From the phase diagram of Fig.4.3 we would like to extract the concentrations of the isotropic-nematic phase boundaries to compare them with the one calculated for the coarse grained model.

The contribution of Figs.4.3 and 4.4 enables us to derive the complete phase diagram. We plotted in Fig.4.3 the line $T_m - \Delta T(T_m)$; this line is the best approximation that we can afford for the boundaries of the nematic region of the coexistence. Between I and N phase boundaries (coloured regions in Fig.4.3) we observed a phase coexistence between isotropic and nematic, while below these regions the phase is fully nematic.

The same procedure was implemented for all the other samples. In Fig.4.5 the $T_m$ is plotted versus the DNA concentration for all samples and at the various concentrations scanned. $\Delta T$ is indicated by vertical lines.

The phase diagram in Fig.4.5 shows that:

- For racemic allAT at 1,8M of NaCl the phase boundaries at 20°C are a little higher than for R-DNA without added salt, the isotropic phase is found at 550 g/l while the nematic at about 650 g/l. For L-allAT the isotropic boundary is about 640 g/l while the nematic is 670 g/l. For the D-allAT we couldn't directly measure the isotropic phase boundary, while the nematic one is about 580 g/l.

- At equal salt concentration D- and L-allAT duplexes are not identical: at equal concentration, the phase of D-allAT has a higher $T_m$ than the phase of L-allAT. Moreover, the width of the coexistence range is different between the two sequences: D-allAT has a wider coexistence region than L-allAT.

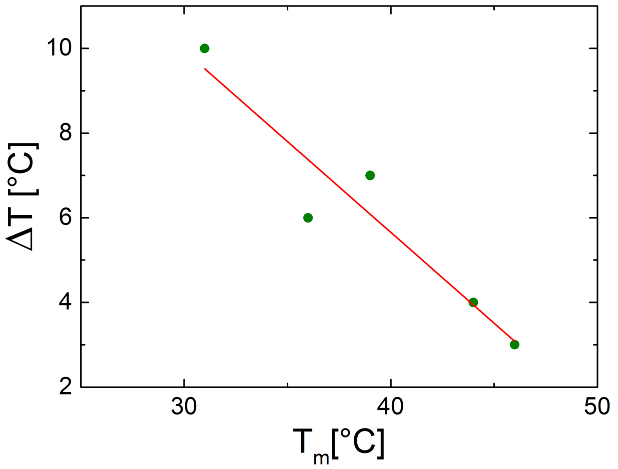


**Figure 4.4:** $\Delta T$ vs $T_m$ for the racemic mixture of AllAT without added salt. The red line is the best linear fit of the data.

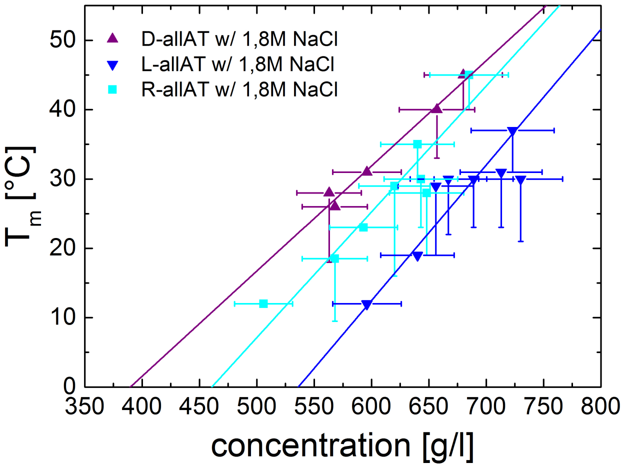


**Figure 4.5:** Plot of the melting temperature of the nematic phase as a function of DNA concentration for racemic (R) allAT at 1,8M of NaCl (light blue squares), D-allAT at 1,8M of NaCl (purple triangles) and L-allAT at 1,8M NaCl (blue triangles). The straight lines are obtained fitting the filled dots for every sequence.

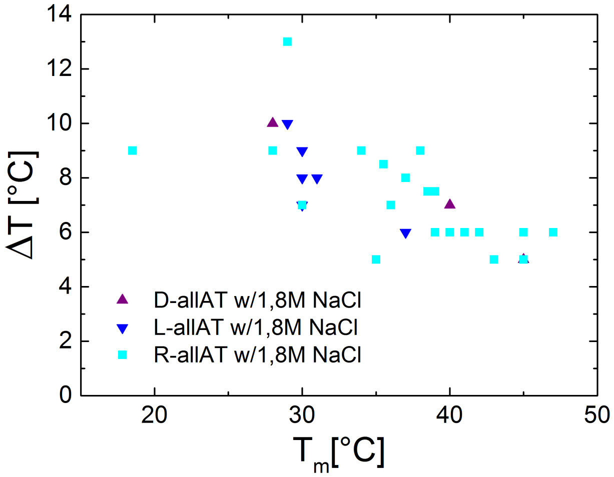


**Figure 4.6:** $\Delta T$ vs $T_m$ for D-AllAT (purple triangles), L-AllAT (blue squares) and racemic R-AllAT (light blue squares). All the solutions are at 1,8 M NaCl.

- Within the experimental errors the racemic phase with 1,8M NaCl behaves like the average between the corresponding D- and L-allAT phases.

- If we increase the salt concentration the melting temperature at the same concentration increase, like we can see for the two racemic samples. This is a reasonable result because the presence of salt screens the negative charges on the helices and then it increases the stability of the phase.

- The D-AllAt and the L-AllAT nematic phases have different melting temperatures at the same concentration. We couldn't find an explanation for this difference because the two sequences were synthesized in the same way and the samples were assembled with the same procedure. Even the salt concentration of both kind of samples was measured a posteriori and no difference was found.

Again, since the concentration measurements were strongly affected by errors, here too we preferred to extract the dependence of $\Delta T$ on $T_m$ rather than on the concentration. We plotted the $\Delta T$ vs $T_m$ for all the samples studied (Fig.4.6). This plot seems to be more noisy then the corresponding one obtained for the R-AllAT without added salt. We could not find the reason for such behaviour but further analysis and measurements will be done.

Using an analogous procedure as the one described for the R-AllAT without added salt, in Fig.4.7 we draw the concentration boundaries of the isotropic-nematic coexistence for all the samples studied.

Overall, a clear feature emerging for all data, despite their large uncertainties, is that the amplitude of all the four coexistence regions seems to get smaller while the concentration is increase. The phase boundaries in Figs.4.3 4.7 and especially those from the racemic mixtures, will be used as a reference to develop the model.

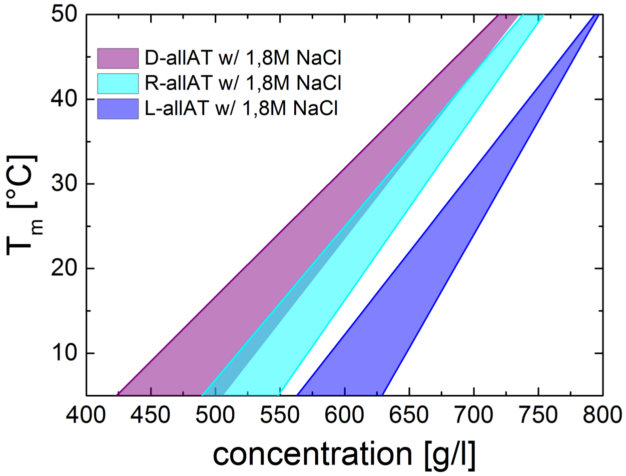


**Figure 4.7:** Measured coexistence regions for racemic (R) allAT at X M of NaCl (green region), racemic (R) allAT at 1,8M of NaCl (light blue region), D-allAT at 1,8M of NaCl (purple region) and L-allAt at 1,8M NaCl (blue region). The vertical lines stand for the $\Delta T$ range of samples whose concentration has been shifted to the fitted one

## 4.4 Order parameter

We also chose to characterize the isotropic-nematic transition by measuring the order parameter $S$ as a function of temperature. A simple and yet effective way to experimentally determine the order parameter is to directly measure the order parameter of a LC phase of DNA, so the best way to obtain an experimental value is to measure the birefringence ($\Delta n$) of the phase (see Fig.4.9) by using a compensator in polarized microscopy. $\Delta n$ is directly proportional to the order parameter, as the 4.1 shows.

$$S = \frac{1}{\gamma}\frac{\Delta n}{c} \tag{4.1}$$

where c is the DNA concentration of the phase in g/l, while $\gamma$ is a constant that varies with the shape and composition of the LC crystal. $\gamma$ is the value of $\Delta n/c$ that would be obtained in a perfectly ordered DNA sample. To obtain $\gamma$ we adopted the reasonable assumption that in the LC columnar phase the linear aggregates are all parallel to each other, so the order parameter can be set to $S = 1$. This assumption is also strengthened by the fact that models developed for the columnar phase of aggregating cylinders [31], $S$ is found to be $S \simeq 1$.

We measured the birefringence and the concentration of some columnar phases of both racemic AllAT samples with different salt concentrations and we determined $\gamma$ for each of them. For the racemic sample without salt added we found an average value of $\gamma = 5,58 \cdot 10^{-5}$ l/g, while for the sample with 1,8M of NaCl $\gamma = 4,95 \cdot 10^{-5}$ l/g. As a cross-check on the T-dependence of $\Delta n$, we also scanned the intensity of the light transmitted through these samples in a microscope between crossed polarizers. While this measurement does not yield the value of $\Delta n$, it is proportional to the birefringence enabling a continuous measurement as a function of the temperature.

In Fig.4.9 some measurements of the birefringence as a function of temperature are shown. Also the intensity measured at different temperatures between crossed polarizers is plotted. Since the intensity is proportional to the birefringence, we adjusted the scale of the intensity to superimpose the absolute values of the two quantities. When the temperature is raised so that the isotropic phase appears in coexistence with the nematic, the measurement of the intensity is unavoidably affected by the presence of the isotropic phase. The growth of the volume fraction of the isotropic phase consequent to the melt of the progressive melt of the nematic phase yield a continuous decrease of the measured intensity in the coexistence region.

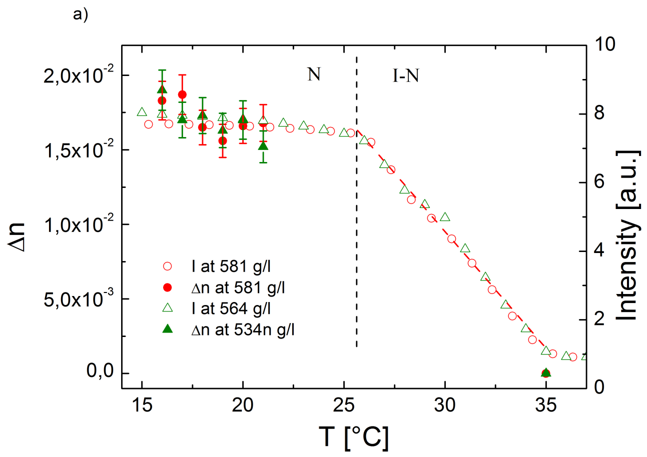


**Figure 4.8:** *
Measured birefringence of the nematic phase at different temperatures and concentrations of racemic allAT with 1,8M of NaCl (a),full symbols)

In Fig.4.10 we plotted the order parameter vs the temperature measured for some racemic samples.

## 4.5 A coarse grained model for the DNA nematic phase

De Michele et al. [28] developed a simplified model to predict some characteristics of the nematic phase of stiff superquadrics from the geometrical properties of the monomers. Specifically, they calculated the isotropic and nematic phase boundaries as a function of the binding free energy between two monomers $\beta\Delta G$, the dimensions of the superquadrics (diameter $D$ and aspect ratio $X = D/L$, where $L$ is the length of the monomer) and the persistence length of the linear aggregates $L_P$ formed by a large number of monomers stuck together and a few other parameters that are fixed by the approximations used to build the model. The monomers have a sticky site at both ends, whose binding energy is modelled as a square potential with a cut off distance, which was chosen to be the typical distance between two consecutive couples of nucleobases in the double helix. Two monomers also interact via steric repulsion.

To apply this model to solutions of AllAT duplexes we set the length of the monomers as similar as possible to the one calculated from the crystallographic reconstruction of the AllAT double helix [32]. $X$ then came to depend only on the diameter $D$. The diameter

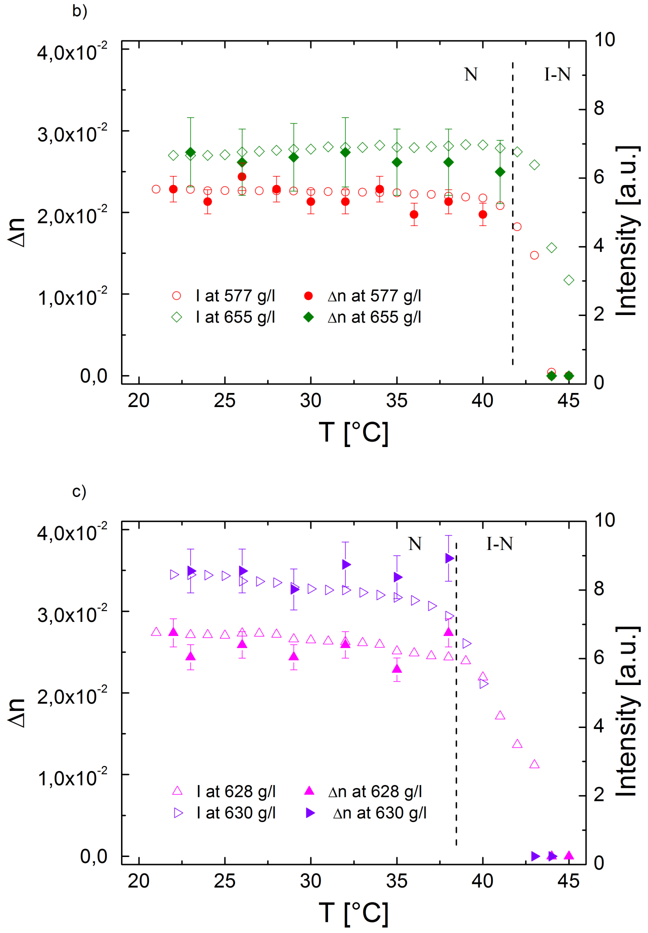


**Figure 4.9:** Measured birefringence of the nematic phase at different temperatures and concentrations of racemic allAT without added salt (b) and c), full symbols). The intensity of the transmitted light between crossed polarizers is shown for the two samples (empty symbols, a), b) and c) ). The black dashed vertical line separates the nematic (on the left of the line) and the isotropic-nematic coexistence region (on the right side of the line). In the isotropic-nematic coexistence region we could not measure the birefringence because the nematic regions were too small, while the intensity was measured in a region that contained both isotropic and nematic domains. As the red dotted line in a) shows, the intensity in the coexistence region decreases linearly with the temperature, as it is expected for a quantity that depends on the volume fraction of the nematic phase.

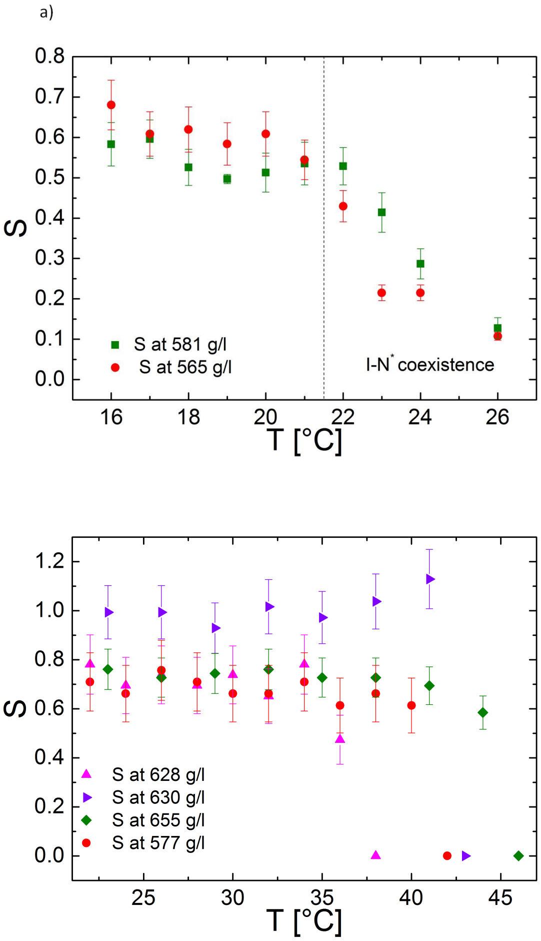


**Figure 4.10:** Order parameter calculated from the birefringence of racemic allAT with 1,8M of NaCl (a),full symbols) and without added salt (b), full symbols).

was set free but within a reasonable range, i.e. between 1,8 and 2 nm. Even if there is a measurement available as for the length of the duplex from crystallography, the diameter was set free because the model does not take into account the electrostatic interactions between the monomers, that are present in a solution of DNA duplexes. Such interactions can be approximated by steric interactions between two molecules whose diameter is different than the physical diameter. All the other parameters were set to best fit the experimental data as explained below. We compared the measured order parameter as a function of temperature and the phase boundaries at room temperature (20°C) to the calculated ones and we found two unique sets of parameters that reproduce both quantities. The first set reproduces the racemic phase without added salt while the second fits the phases obtained adding 1,8M of NaCl. The we varied the strength of the binding energy between the ends of two monomers (i.e. the depth of the square well potential mimicking the stacking interaction) as a function of the temperature and we tried to reproduce the measured coexistence region of the racemic samples, as explained in the following.

### 4.5.1 Phase boundaries at 20°C

We firstly sought for a set of parameters to be put in the coarse grained model in such a way that the phase boundaries measured at 20°C could be reproduced.

Within the coexistence region the system separates in nematic and isotropic domains with two different concentrations but the same osmotic pressure and chemical potential. Both quantities by definition depend on the free energy of the system, as shown in 4.2:

$$\mu = \frac{\partial F}{\partial \phi} \nu \pi = -\frac{1}{\nu}\left(F\nu - \phi\nu\right) \tag{4.2}$$

To calculate $\mu$ and $\pi$ from the model in a range of concentrations that comprehend the experimental phase boundaries we firstly evaluated the free energy $F_I$ and $F_N$ which are explicitly given by the model as a function of the parameter listed before. In this analysis we kept fixed the length of the duplex ($L$) and thus its axial ratio($X$) using the value obtained from the crystallographic reconstruction of a AllAT duplex [32], and we explored different predictions of the model by changing $\beta\Delta G$ and $L_p$ as described in the following.

To calculate the isotropic-nematic phase boundaries we implemented the following procedure:

- we arbitrarily chose a set of parameters (the binding free energy $\beta\Delta G$, the persistence length $L_P$, the diameter of the superquadrics $D$ and its axial ratio $X$).
- We studied $F_I$ that depends on the parameters listed before and on the average length of the linear aggregates ($M$), and $F_N$ that also depends on $\alpha$. $\alpha$ is a parameter proportional to the order parameter of the nematic phase.
- At the minimum of the free energy $M$ can be written explicitly both for isotropic and nematic phases. We inserted these expressions in $F_I$ and $F_N$.
- We calculated the value of $\alpha$ that minimizes $F_N$ and we inserted this value in $F_N$.
- We calculated $\mu$ and $\pi$ for the nematic and isotropic phases and we searched for the pair of concentrations ($c_I$ and $c_N$), one for the isotropic and one for the nematic phase, that yield to equal $\mu$ and $\pi$ between the two phases.

- We compared the pair of concentrations found to the experimental coexistence region of both racemic samples measured at 20°C. If the two pairs of concentrations were not compatible we changed the parameters and we repeated the process from the beginning.

Through these procedure we obtained a set of parameters that yielded to a couple ($c_I$, $c_N$) compatible to the measured one for the racemic samples. As anticipated, we chose to perform a double check on these parameters by using them to fit the measured order parameter.

### 4.5.2 Comparison between the measured order parameter and the one extracted from the model.

We calculated the order parameters of the nematic phase of the superquadrics as a function of the temperature through the coarse grained model to compare it with the experimental measurements. The order parameter is a function of $\alpha$, as the 4.3 shows.

$$S = \frac{1}{3\alpha} \tag{4.3}$$

At every fixed temperature we calculated $\alpha$ using the nematic free energy. To compare the calculated $S$ to the measured one, we chose a set of measurements of Fig.4.10, we determined the concentration of the sample used for these measurements and we used this concentration as the monomer concentration in the coarse grained model. We also fixed the value of the binding energy $\beta\Delta G$, the persistence length $L_p$,the diameter and the axial ratio of the duplex using the values that best fit the coexistence region at 20°. We then calculated $F_N$ that, as explained before, depends on $M$ and $\alpha$. As already done for the determination of the concentrations at the isotropic-nematic phase boundaries, we substituted into $F_N$ the analytic expression for $M$ that minimizes the free energy, so that the free energy became just a function of $\alpha$. Then we minimized the free energy as a function of $\alpha$ in order to find the value of $\alpha$ and as a consequence of S at the thermodynamic equilibrium.

This procedure was done at fixed temperature. To vary the temperature we changed the binding energy, because the two quantities are connected by the definition of free energy itself.

Fig.4.11 shows that the calculated order parameter has a similar behaviour to the experimental one until the phase starts to melt. S decreases of about 10% while the sample remains fully nematic. In other self assembly induced LC, as the chromonics, a similar phenomenon has been observed [38] [39] [40]: the order parameter is seen to have a small decrease in a long range of temperatures, and it has a sudden decrease near the melting temperature of the phase. On the other hand in thermotropic LC, in which the order is not due to self assembly, the order parameter has a different behaviour: it starts decreasing as soon as the temperature is raised, till it reaches the zero when the phase is melted [41][42][43][44]. The behaviour obserevd in chromonics and LC of short DNA duplexes indicates that the LC formation by aggregation is a more discontinuous process than the isotropic-nematic transition in thermotropics, because it involves two phenomena at the same time: the increase of orientational order and also the increase of the aggregation of the monomers.

Overall the model well predicts the behaviour of the measured order parameter, and the aggregates remain quite long for a large range of temperatures until the thermal energy becomes of the order of the binding energy of two rods.

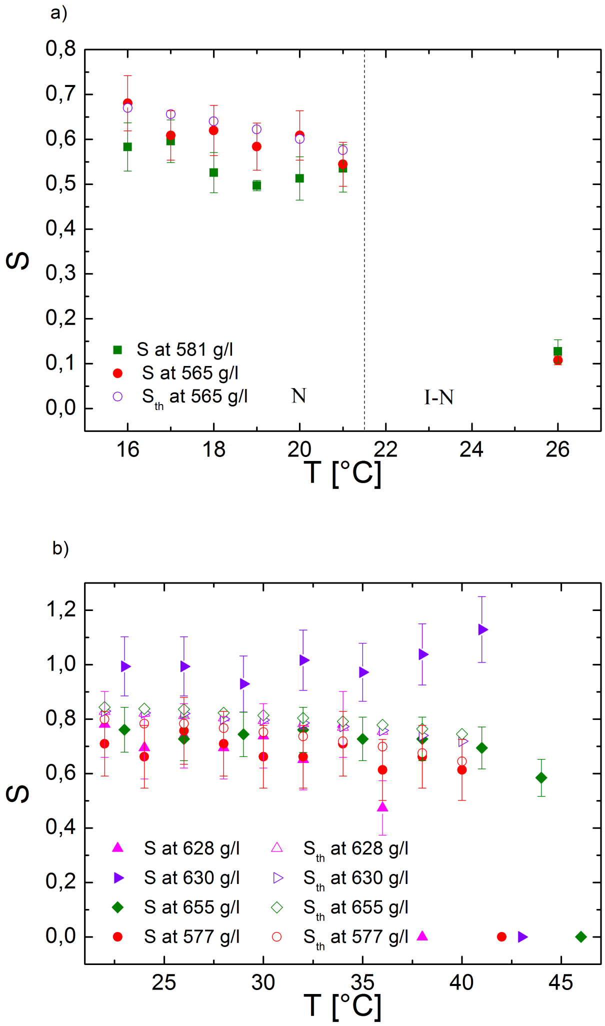


**Figure 4.11:** Measured (full symbols) and calculated (empty symbols) order parameters as a function of temperature for the racemic allAT with 1,8M of NaCl a) and without added salt b). The dotted line indicates the temperature at which the melting of the phase starts.

### 4.5.3 The best set of parameters to match the coexistence region at room temperature

We tried so far to find the best set of parameters (the binding free energy $\beta\Delta G$, the persistence length $L_P$, the diameter of the superquadric $D$ and its aspect ratio $X$) that could be set into the coarse grained model developed by De Michele et al. in order to predict the formation of a nematic phase that:

- shows a isotropic-nematic phase boundaries whose concentrations are compatible with the ones experimentally measured of a racemic mixtures of AllAT duplexes at 20°C.

- These parameters should also predict a behaviour of the order parameter at 20°C compatible to the measured behaviour in a nematic phase of racemic AllAT.

We looked for two sets of parameters, one that reproduced the above listed properties of the racemic mixtures of AllAT without added salt, and the other that reproduces the properties of the racemic mixture with 1,8M of NaCl. added. At the beginning of the procedure we fixed the length of the monomers so that the aspect ratio would depend only on the diameter. Then we changed the other three parameters in order to find the best set that could reproduce the experimental data at 20°C. We found just one set corresponding to the racemic without salt added and another set for the racemic samples at 1,8M NaCl. These parameters are listed in table4.1.

| | $l_p$ | $\beta\Delta$G | D [nm] | X | $\phi_I$ [g/l] | $\phi_N$ [g/l] |
|---|---|---|---|---|---|---|
| R-allAT | 7,5 | -1,24 | 1,96 | 2 | 503 | 588 |
| R-allAT w/ 1,8M NaCl | 4,7 | -3,84 | 1,86 | 2 | 585 | 652 |

**Table 4.1:** Best sets of parameters that fit the experimental data (the I-N phase boundaries at 20°C and the order parameter) for the two racemic samples. On the right there are the concentration boundaries obtained with the listed parameters.

The values obtained for the two samples are different: the binding energy is stronger for the sample that has a higher salt concentration. This result is reasonable because the strength of the stacking interaction is known to increase with the salt concentration within the double helix and in coaxial stacking [7]. We can also notice that the helix diameter is smaller for the second sample. It can be related to the difference between the rods and the duplexes. The latter are charged, while the former are not charged, so the only way to compare the nematic phases made by these two molecules is to tune the diameter of the rods. Stronger the interactions bigger should be the diameter of the rods to better reproduce the experimental nematic phases.

The free energy $\beta aG$ that was used as a free parameter in the coarse grained model consiste of two basic contributions: the free energy of bonding and the entropic term due to the variation of the conformational space of a duplex when bonded to another duplex. This latter term is proportional to the ratio between the volume accessible to the center of mass of a duplex while interacting with another duplex (called bonding volume $v_b$) and the volume accessible for the center of mass without interactions at play. The free energy of bonding instead consists of an energetic and an entropic contribution. The energy gain of the interaction depends the depth of the potential well that describes the interaction, while the entropic term depends of the decrease of the conformational space of both strands that form the bond.

### 4.5.4 Best set of parameters for the temperature dependent phase boundaries

Once obtained a set of parameters that well fits the phase boundaries at 20°C and the order parameter as a function of the temperature of the two racemic samples, we tried to understand if the same sets of parameters could reproduce the measured I-N phase boundaries as a function of the temperature. Since temperature and free energy enter the model via the relative $\Delta G/T$, we explored the phase boundaries by holding fixed all the parameters except for $\beta\Delta G$, that was varied in a range that yielded concentrations of the phase boundaries most similar similar to the measured ones. The result is shown in Fig.4.12, open dots.

As appears from the figure, for both racemic samples the values of the concentration boundaries are not compatible with the experimental ones because the ratio $\frac{c_N}{c_I}$ increases when the concentration increases, while the experimental ratio decreases. An obvious problem of this approach is having held $L_P$ as a constant at all the temperatures. This is a direct consequence of the form of the attractive potential well: the model simplifies the interactions between the ends of two monomers using a square well potential, so the persistence length of linear aggregates of monomers is found to be constant when the temperature (i.e. the depth of the well) is varied. This is an oversimplified description of the interaction between two duplexes because of two reasons. The first is that the attractive potential between two duplexes is a soft potential, continuously decaying with the distance between the ends of the duplexes. Such shape of the potential will provide a persistence length that varies with the temperature: within an aggregate of monomers, the persistence length depends on the average bend between two consecutive monomers. Different regions of the potential well yield different bend angles between two monomers. At fixed temperature the mean bend angle is calculated as an average over all the profile of the potential well, with each region of the potential populated with a probability that depends on the temperature of the system. If the temperature is changed, this probability changes, making accessible regions of highest potential, typically associated to a larger distance, that allows for larger bend angles. As a consequence when the temperature changes, the average bend angle and also the $L_P$ change. The second reason for assuming that the persistence length should vary with the temperature is that the amplitude itself of the potential is expected to change as a function of the temperature. This is indeed known to happen for the stacking interactions in which the entropic terms are known to play a large role [7]. Quite clearly, if the interaction energy weakens upon increasing temperature, the flexibility of the aggregates further increases.

We calculated again the concentrations at the I-N phase boundaries and the order parameter at different temperatures (i.e. different $\beta\Delta G$) using the same coarse grained model, but arbitrarily decreasing the persistence length while increasing $\beta\Delta G$. In Fig.4.12 the results are shown. The values of the parameters used in the model are listed in table 4.2. The comparison between the experimental and the calculated phase boundaries clearly indicates that the variability of $L_P$ on the temperature yield a better compatibility between data and results of the model. In fact decreasing the $L_P$ while increasing $\beta\Delta G$ provides that the amplitude of the coexistence region gets smaller when the concentration is increased.

The square well model used here does not provide any insight into the temperature dependence of the potential. Indeed, even by adopting the results of Table 4.2, as accurate the square well model doe not enable extracting information about the stacking interactions between two duplexes. A better physical description of the stacking interaction is the goal of the studies explained in the next section.

From table 4.2 the value of the free energy $\Delta G$ associated to a stacking event between

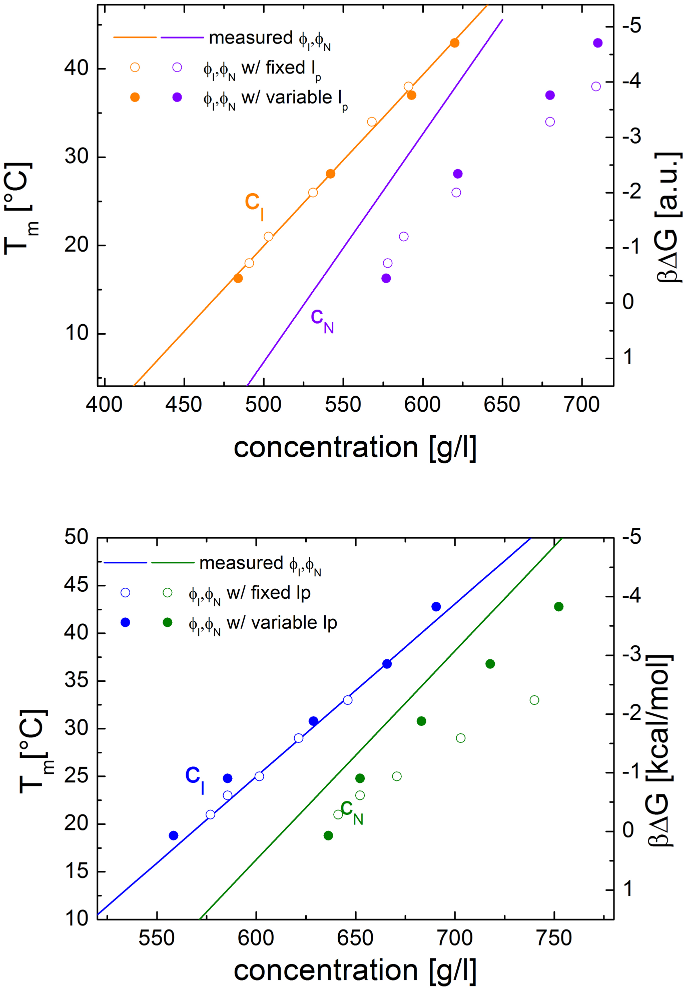


**Figure 4.12:** Comparison between the concentration phase boundaries measured and calculated using the coarse grained model. The straight lines are the measured concentrations of the isotropic-nematic phase boundaries. The empty dots are the phase boundaries calculated from the model using a fixed persistence length for every value of binding energy, while the full dots are calculated with the same values of the parameters but with a variable persistence length. On top the data for R-AllAT are shown, while on bottom there are the data for R-AllAT with 1,8M NaCl.

| R-allAT | | | R-allAT w/ 1,8M NaCl | | |
|---|---|---|---|---|---|
| T | $l_p$ | $\beta\Delta$G | T | $l_p$ | $\beta\Delta$G |
| 17 | 8,4 | -0,34 | 18 | 6 | 0,16 |
| 20 | 7,5 | -1,24 | 25 | 4,7 | -0,84 |
| 28 | 6 | -2,44 | 31 | 3,3 | -1,84 |
| 37 | 4,2 | -3,94 | 37 | 2,7 | -2,84 |
| 43 | 3 | -4,94 | 42 | 2,5 | -3,84 |

**Table 4.2:** Best set of parameters used to reproduce the experimental measurements with the coarse grained model. The persistence length is variable.

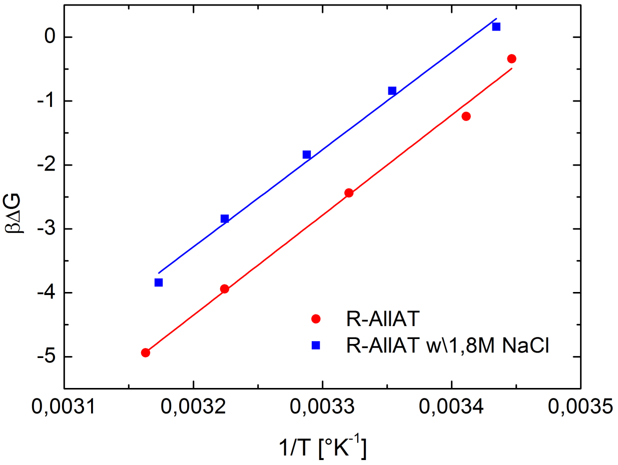


**Figure 4.13:** $\beta\Delta G$ as a function of the inverse of the temperature for R-AllAT and R-AllAT with 1,8M NaCl.

two duplexes at 37°C is about -1,74 kcal/mol, that is compatible with the value obtained for nicked duplexes [7]. In Fig.4.13 the $\beta\Delta G$ is shown as a function of the inverse of the temperature. Form this plot we calculated the enthalpic term by fitting the behaviour with a straight line. The value obtained is about 30 kcal/mol, an order of magnitude larger than the one measured within the nicked DNA. The entropic terms are difficult to compare because they are determined by different phenomena: the entropic contribution of the free energy measured with the nicked sequences of DNA depends only on the decrease of the number of conformations of the two strands after the formation of the duplex. Instead the entropic contribution obtained from the coarse grained model depends also on the bonding volume, that keeps into account the loss of conformational space of one duplex due to the interaction with another one.

## 4.6 Determination of the shape and strength of the stacking interactions through combining atomistic simulations and the coarse grained model

Using the coarse grained model developed by De Michele et al. we found two sets of parameters that yield to the best possible compatibility between the experimentally measured and calculated concentration of the isotropic-nematic phase boundaries, and also between the measured and calculated order parameter at different temperatures. One set of parameters was chosen to best reproduce the properties of the nematic phase of R-AllAT without added salt, while the other provided the best compatibility with the properties of the nematic phase of R-AllAT with 1,8M NaCl.

To gain new insight into the meaning of these parameters we compared them to an atomistic model of two duplexes of AllAT that interact at the terminals at the same ionic strength used in the experiments (1,8 M of NaCl). By using all-atom simulations we investigated the phenomena controlling the temperature dependence of the persistence length of the linear aggregates. The work described in this chapter was organized as follows:

- to understand if the shape of the potential could yield a $L_P$ that varies with the temperature, we simulated via molecular dynamics the interactions of two duplexes of AllAT at various temperatures and extracted the average bend angle between the principal axes of the duplexes. From this we could calculate the persistence length of the aggregates as a function of the temperature.

- As a second approach we adopted the shape of the potential resulting from the atomistic simulation but used its overall amplitude as a variable quantity. To this aim we implemented the following procedure:
  - we reconstructed the 3D potential well between the ends of two duplexes from the projections of the potential along suitable coordinates;
  - we calculated on the basis of such potential mean force the average bend angle between two interacting duplexes at a given temperature, i.e. at a fixed depth of the potential well;
  - we changed the depth of the potential well and we calculated the corresponding persistence length.

- We then tried to reproduce the binding free energy and the persistence length used in the coarse grained model to fit the data measured with the R-AllAT with 1,8M NaCl at the different temperatures sampled. Both the energetic and entropic terms were varied as the temperature was changed.

The general concept behind this approach is the following: the coarse grained model enables determining parameters that cannot be easily related to the actual stacking interaction. The all-atom simulation by itself yields a weak temperature dependence of $L_P$, too weak to account for the observations. This might be due to intrinsic overestimation of the stacking interactions, well documented in the literature. It might also be due to an approximation that was used to perform the simulations, as described later. We thus adopted the shape of the stacking interaction as obtained by all-atom simulations and used its amplitude as a free variable, to be able to match the parameters in table 4.2. This leads to a determination of the temperature dependence of the stacking interaction.

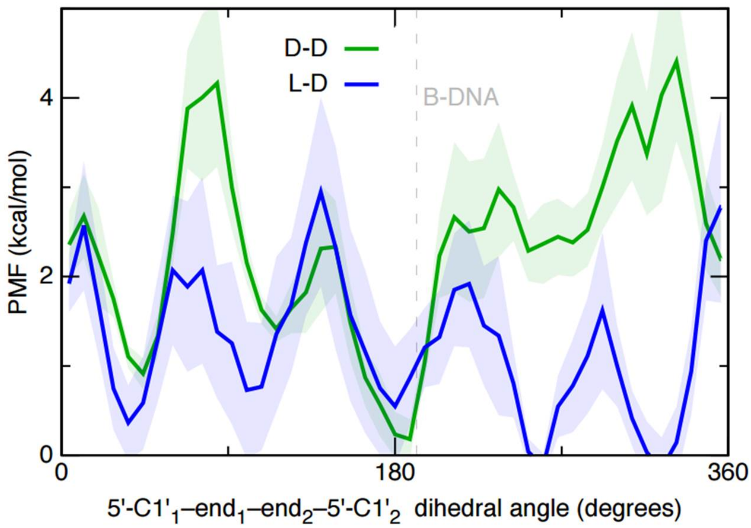


**Figure 4.14:** PMF vs azimuthal angle for D-D and D-L configurations, at 20°.

For the purposes listed above we started a collaboration with Maffeo and Aksimentiev. They had already made some all atom simulations on a system of two blunt ended duplexes of one sequence called Dickerson Dodecamer ('DD' sequence) at 100 mM NaCl. It was found that the conformation of the terminal of the duplex determined the reciprocal position of the two molecules when they are interacting. They defined the azimuthal angle as the angle between the projections of the vectors connecting the O5′ and O3′ atoms of the terminal base pairs into the plane normal to the common DNA axis. If the two duplexes are kept with their principal axes aligned along the same direction, the different azimuthal angles are assumed with different probabilities. For example for two DD duplexes it was found that they preferentially assume a reciprocal orientation such that the azimuthal angle is about 180°, while if an extra phosphate group is added at the terminals of both duplexes, the favorite azimuthal angle is about 20° [29]. As a reference, the azimuthal angle corresponding to a configuration of helical continuity is 36°.

As a first step, an analogous study was done for the AllAT duplexes. In this case the restraint on the axes of the duplexes was not applied because of the high ionic strength of the system. In this condition and for such simulations the probability of opening and closure of the bases is extremely low, so the restraint is not necessary. Different azimuthal angles were sampled by harmonically restraining the reciprocal orientation of the two duplexes around the chosen angle and it was found that the interaction strength (i.e. the average value of the potential mean force, PMF) depends on the azimuthal angle. In Fig.4.14 we plot the PMF as a function of the azimuthal angle calculated from the simulations of two AllAT interacting duplexes.

The simulations discussed here and below were made considering two situations, one where the two interacting duplexes had the same chirality, and the other configuration where the two duplexes had opposite chirality (L-D duplexes). For the first set up we considered two right-handed duplexes (D-D duplexes), and we assumed that the results should be identical to the one we would obtain using two left-handed duplexes because these molecules are symmetrical. From Fig.4.14 we noticed that the minima of the PMF are found at different azimuthal angles for the two systems. For the D-D duplexes the minimum of the PMF is assumed at an angle of about 180° as found before for Dickerson Dodecamer duplexes, while L-D duplexes have a more funneled PMF landscape. They show two minima for the PMF at azimuthal angles of about 250° and 330°.

### 4.6.1 Persistence length of AllAT linear aggregates

After obtaining the PMF as a function of the azimuthal angle ($\phi$), we calculated the average bend angle ($\theta$) between two interacting duplexes at different temperatures. The simulations were similar to the previous ones: we considered both D-D and L-D systems, and for every considered temperature we sampled all the azimuthal angles. We defined the bend angle between two duplexes as the angle between their principal axes. Specifically, we considered the angle formed by the two segments AB and BC, where A is the center of mass of one duplex, B is and the center of mass of the two base pairs at the junction between the two duplexes, and C is the center of mass of the other duplex. No restraint on the hydrogen bonds between the bases was applied, with the result that for each duplex at high temperature the last couples of bases away from the interaction region break the hydrogen bonds and open. Since this does not affect the couples of bases that take part into the stacking interaction between the other duplex we did not add any constraint. As the simulation was running, different bend angles were assumed by the duplexes. At every configuration of azimuthal and bend angles we computed the PMF. From the simulations we extracted the average PMF and then the probability of that configuration as $P(\theta, \phi) = exp(-PMF/k_B T)$ a function of the D-D or L-D configurations, the azimuthal and bend angles and the temperature $T$ .

In Fig.4.15 there is a 2-D map of the PMF as a function of the bend angle and the azimuthal angle, for two interacting D-D or two L-D duplexes at 20°C. For every sampled temperature many simulations were run with identical replica of the same system, then they were averaged to assure a good sampling of the possible conformations of the duplexes. For a good sampling, the azimuthal angle was varied from 0° to 360° with a step of 4°. What we obtained is then the probability of the bend angle as a function of temperature and azimuthal angle.

The average bend angles $\theta_{DD}$ and $\theta_{LD}$ for respectively D-D and L-D configuration at fixed temperature are calculated as an average over all the azimuthal angles. Equation 4.4 express the mean bend angle for D-D configuration, but an analogous formula can be written for L-D duplexes.

$$\theta_{DD} = \frac{\sum_{\phi=0,360^\circ} \theta(\phi) P(\theta(\phi))}{\sum_{\phi=0,360^\circ} P(\theta(\phi))} \tag{4.4}$$

The average between the two configuration is calculated as the weighted mean for every temperature sampled. If $PMF_{DD}$ and $PMF_{LD}$ are the average potential mean force of the D-D and L-D configurations at fixed temperature, the average bend angle becomes:

$$\theta_M = \frac{e^{\beta PMF_{DD}} \theta_{DD} + e^{\beta PMF_{LD}} \theta_{LD}}{e^{\beta PMF_{DD}} + e^{\beta PMF_{LD}}} \tag{4.5}$$

From the average bend angle we calculated the persistence length of linear aggregates of AllAT duplexes. If we roughly approximate such aggregates to a freely rotating

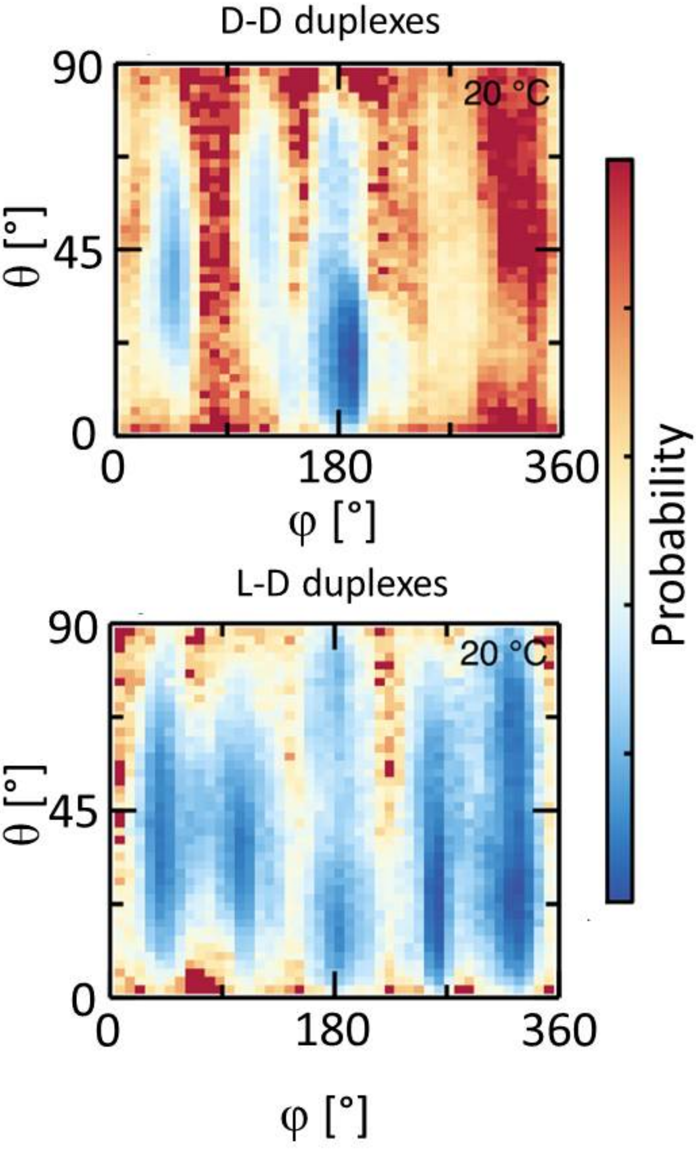


**Figure 4.15:** Bend angle $\theta$ calculated from the atomistic simulations at 20°C for two D-D and L-D duplexes as a function of the azimuthal angle $\phi$. From red to blue the PMF decreases and the probability increases. The probabilities of both D-D and L-D configurations are normalized such that the sum of the probabilities over both the different bend and the azimuthal angles are equal to 1.

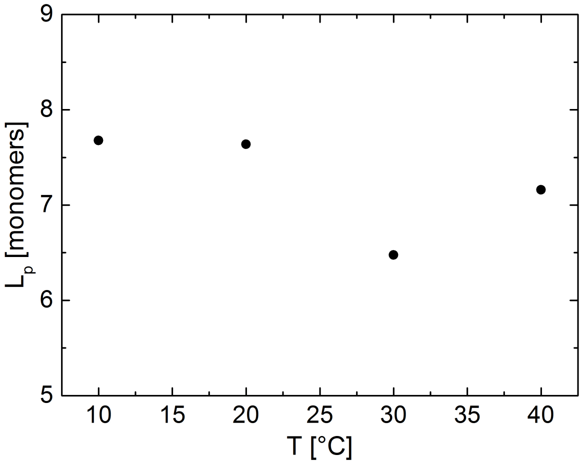


**Figure 4.16:** Persistence length $L_p$ averaged over D-D and L-D configurations at different temperatures. It is expressed in monomer units.

chain with a fixed bend angle ($\theta_M$) between the axes of consecutive monomers, the persistence length depends on the bend angle as in equation 4.6.

$$L_p = \frac{1}{1 - cos(\theta_M)} \tag{4.6}$$

In table 4.3 we report the average bend angle and the persistence length as a function of the temperature and the configuration. We can notice that the L-D configuration allows a broader range of bend angles than D-D setup, as it is shown also in Fig.4.15.

| D-D duplexes | | | L-D duplexes | | |
|---|---|---|---|---|---|
| T [°] | $\theta_{DD}$ [°] | $L_p$ [monomers] | T [°] | $\theta_{LD}$ [°] | $L_p$ [monomers] |
| 10 | 28.5 | 8.3 | 10 | 30.8 | 7.1 |
| 20 | 26.8 | 9.3 | 20 | 33.7 | 6.0 |
| 30 | 31.7 | 6.7 | 30 | 32.9 | 6.2 |
| 40 | 29.1 | 7.9 | 40 | 32.5 | 6.4 |

**Table 4.3:** Average bend angles and persistence lengths calculated from atomistic simulations at different temperatures and for the two D-D and L-D configurations.

In Fig.4.16 the persistence length averaged over D-D and L-D configuration is shown as a function of the temperature sampled. The average was calculated as the algebraic mean. We noticed that though the value at about 20°C is similar to the value required from the coarse grained model to reproduce the experimental phase diagram (7,5 monomers), the persistence length did not change while increasing the temperature. This fact is not acceptable on the basis of the experimental measurements so we can infer that a soft attractive potential with a fixed depth between the ends of two duplexes is not enough to assure a dependence of $L_P$ on temperature.

Because we could not find an agreement between the experimental data and the results of the simulations, we tried a different approach. We reconstructed the 3D shape of the potential well between the two duplexes, then we varied the depth of the well as a function of the temperature in order to see if there was any change in the persistence length such as to reproduce the values adopted in the coarse grained model.

### 4.6.2 Modeling of the potential well between two interacting duplexes

To reconstruct the 3D shape of the potential well, we calculated the PMF between two interacting AllAT duplexes in two distinct configurations.

For the first configuration we fixed the orientation of the duplexes along the same axis in such a way that they were coaxially aligned. Then we varied the end to end distance between the duplexes and we calculated the corresponding PMF. Since there is no obvious reference system, we adopted the direction of the end to end distance to be the average of the principal component vectors corresponding to the smallest moment of inertia of each of the duplexes. We performed such simulations for both D-D and L-D configurations. The azimuthal angle between the duplexes was restrained to the value corresponding to the minimum of the PMF calculated before (see Fig.4.14, for DD configuration we fixed $\phi \simeq 180°$, while for L-D $\phi \simeq 260°$.) and the hydrogen bonds between the bases were reinforced with harmonic restraints in order to avoid the opening of the terminal couples. The simulations were made at 10°C. In Fig.4.17 the strength of the interaction energy ($E_{DD}$ and $E_{LD}$ respectively for interactions between D-D and L-D duplexes) between the two duplexes vs the vertical distance (z) is shown. The interaction potential E is obtained from the PMF by a simple translation. We supposed that at the minimum of the PMF corresponds a minimum in the interaction strength, and at the plateau of the PMF corresponds to an analogous plateau of the interaction strength whose value is equal to 0. The z direction is chosen to be the end to end distance between the two duplexes. From this plot we can deduce the depth of the vertical potential well to be around 7 kcal/mol and we can notice that the vertical distance at which the E is minimum is about 3,7 $\AA$, that is higher than the distance between two consecutive couples of bases within a B-form of a double helix of DNA (i.e. 3,2 $\AA$). It is interesting to notice that such large distance is found despite the fact that the terminal bases are constrained to be closed, a condition that certainly favours a closer distance. It is curious that the z distance takes also values smaller than 3,2 $\AA$ because of the thermal fluctuations of the positions of the duplexes. The end to end distances for which the potential is minimum are different for D-D and L-D configurations. The fact that the two configurations have different equilibrium conditions had already emerged from the study of the PMF vs the azimuthal angle (see Fig.4.14). The black and blue profiles ($E_{10DD}$ and $E_{10LD}$) of Fig.4.17 are calculated from the all-atom simulations at the azimuthal angles at which the PMF showed to have a minimum. This is not the only configuration that the system visits, every azimuthal angle can be adopted by the system with a probability that depends on the temperature. We averaged the PMF plotted in Fig. over all the azimuthal angles to obtain the mean value of the PMF at 10°C. The procedure was repeated both for D-D and L-D configuration. We adopted these values as the depths of the potential well at 10°C for respectively D-D and L-D configurations. In Fig.4.17 we plotted also the potential well obtained using the depth averaged over the azimuthal angles.

We made an average of the E ($E_{10}(z)$) over the possible combinations of duplex chirality and over the azimuthal angle at 10°C as shown in equation 4.7 and we obtained

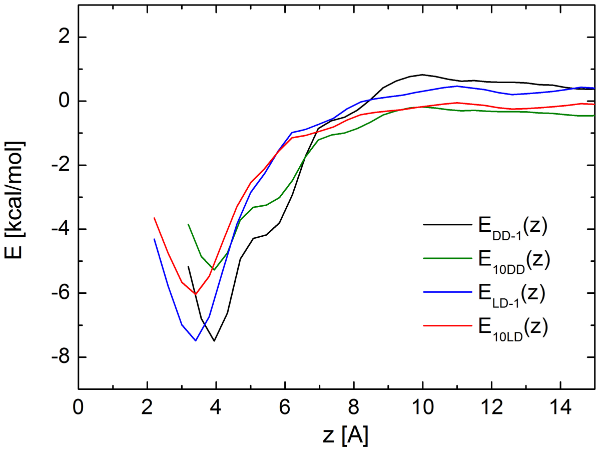


**Figure 4.17:** E as a function of the end to end distance between the two duplexes. $E_{DD}$ and $E_{LD}$ are the potentials of the interaction between respectively D-D and L-D duplexes at $\phi \simeq 180°$ for D-D and $\phi \simeq 260°$ for L-D. $E_{10DD}$ and $E_{10LD}$ are the potentials whose minima are normalized to the values at 10° averaged over all the azimuthal angle, obtained from the data plotted din Fig.4.14.

the profile of figure 4.18.

$$E_{10}(z) = \frac{e^{\beta E_{10DD}(z)} E_{10DD}(z) + e^{\beta E_{10LD}(z)} E_{10LD}(z)}{e^{\beta E_{10DD}(z)} + e^{\beta E_{10LD}(z)}} \tag{4.7}$$

In order to be able to extract useful information from the potential, it is necessary to turn the numeric potential profile onto a simple analytical shape. To this aim we fitted the potential with a parabola for $3,2\,\AA \leq z < 4,7\,\AA$ crossing over a straight line for $4,7\,\AA \leq z < 8,1\,\AA$. The fitting equations are:

$$\begin{aligned} E(z) &= A + Bz + Cz^2 \\ E(z) &= D + Fz \end{aligned} \tag{4.8}$$

where the values of the parameters are $A = 46,42 \pm 0,98$, $B = -24,07 \pm 0,51$, $C = 2,79 \pm 0,06$, $D = -11,43$, $F = 1,38 \pm 0,03$.

For the second configuration we fixed the vertical distance between the ends at the value corresponding to the minimum of the energy $E$ (about $4\,\AA$) and we kept the coaxial orientation of the duplexes frozen, but we increased the distance between the duplexes shearing one of them in the horizontal (xy) plane. We chose to fix the position of one duplex and to shear the other one along two orthogonal directions, x and y, that are related to the fixed duplex. y stands for the direction that links the C1 carbon atoms of the terminal couple of bases, while x is the perpendicular direction that points to the minor groove. For every x and y position, we sampled all the possible azimuthal angles ($\phi$), and for each of them the average interaction energy was extracted from the simulation. From the energy we calculated the probability of the configuration at fixed x (or y) and $\phi$, then we averaged the energy over all the azimuthal angles at a fixed

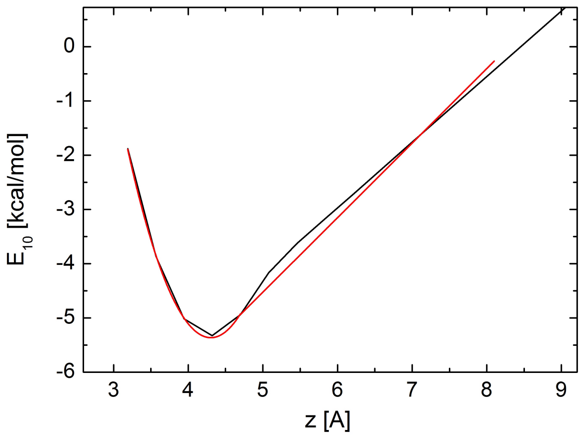


**Figure 4.18:** Average smoothed energy E vs end to end distance between two duplexes. The fitting function is a parabola for $3,2\,\AA \leq z < 4,7\,\AA$ and a straight line for $4,7\,\AA \leq z < 8,1\,\AA$

shearing distance, as shown in formula 4.9. The average energy along y direction is calculated with an analogous formula.

$$E_{10}(x) = \frac{e^{\beta E_{10DD}(x)} E_{10DD}(x) + e^{\beta E_{10LD}(x)} E_{10LD}(x)}{e^{\beta E_{10DD}(x)} + e^{\beta E_{10LD}(x)}} \tag{4.9}$$

In Fig.4.19 the energy $E$ as function of the shear distance is shown for both D-D and L-D configurations. It is normalized in the range $-1$ - $0$ because it was used to modulate the potential extracted as a function of the end to end distance. The procedure is explained below in the text.

We then averaged the $E$ extracted from shearing one duplex from the other to obtain a rough but more general behaviour for the potential mean force as a function of the distance from the center of the terminal base pair of the fixed duplex along the horizontal (x-y) plane. We weighted both $E_{10}(x)$ and $E_{10}(y)$ with their corresponding Boltzmann probability:

$$E_{10}(x|x=y) == \frac{exp(\beta E_{10}(x))E_{10}(x) + exp(\beta E_{10}(y))E_{10}(y)}{exp(\beta E_{10}(x)) + exp(\beta E_{10}(y))} \tag{4.10}$$

We plotted the average $E$ ($E_{10}(x)$) as a function of the shear distance in Fig.4.20. We fitted this behaviour with a parabolic function with the origin in the point of coordinates (0,0). The equation of the fitting function is

$$E(r) = mr^2 + 1 \tag{4.11}$$

where $r = \sqrt{x^2 + y^2}$ and $m = 0,108 \pm 0,002$.

After obtaining the shape of the interaction energy as a function of the shear distance between the duplexes and the end to end distance between the terminals, we reconstructed the shape of the 3D potential well between the duplexes, in such a way to satisfy these conditions:

- the depth of the potential well must be the one calculated from the energy vs the end to end distance at the z coordinate corresponding to the minimum of the energy.

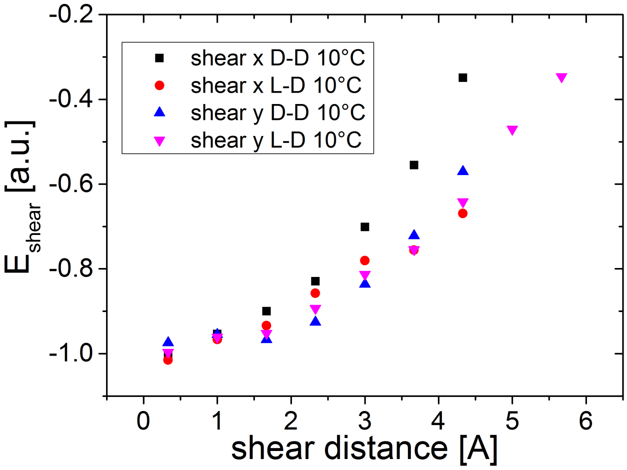


**Figure 4.19:** Normalized $E$ vs x and y shear directions at 10°C for D-D and L-D configurations, at 10°.

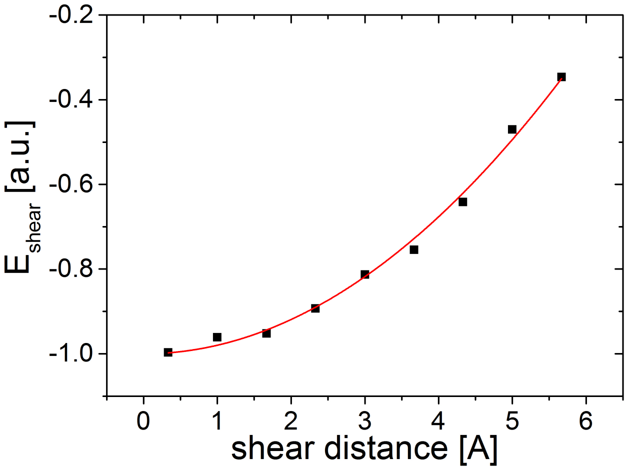


**Figure 4.20:** Normalized average interaction potential vs the shear direction at 10°C. The mean is calculated over the configurations and the sampled shear directions.

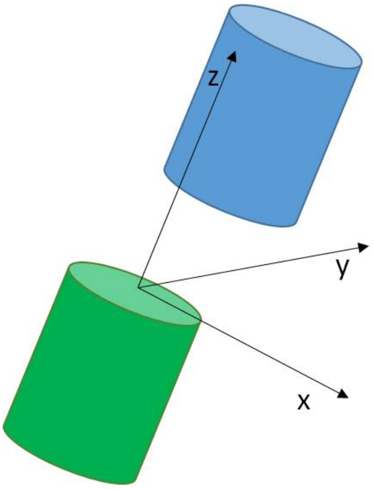


**Figure 4.21:** Scheme of the frame of reference adopted for the reconstruction of the 3D potential well.

- At fixed vertical distance $z_0$ the shape of the potential must have a minimum at zero shear whose depth is the value of the energy vs z at $z = z_0$. To follow this criterion we must properly normalize the shear energy in such a way that the depth of the shear potential is unitary.

We described the 3D potential as the product of the vertical and shear interaction energies:

$$E(r, z) = E_{10}(z)E(r) \tag{4.12}$$

For $3,2 < z < 4,7\,\mathring{A}$ it becomes $E(r, z) = (A + Bz + Cz^2)(G + Hr^2)$ where the reference frame is centred in the center of the terminal couple of bases (the one closest to the other duplex) of one duplex, z is vertical projection of the distance between the centres of the two closest terminal bases belonging to the different duplexes, and x and y are the coordinates along the plane of the couple of bases at $z = 0$. With the same reference frame, for $z > 4,7\,\mathring{A}$ the 3D $E$ can be calculated as $E(r) = (D + Fz)(G + Hr^2)$.

From the shape of the potential well we calculated the average bend angle between two interacting duplexes as a function of the depth of the well and the temperature. This was done in two steps: we calculated the average bend angle $\psi$ for every reciprocal position of the centres of the terminal couples of bases of the two duplexes, then we calculated the bend angle $\psi_{bend}$ averaged over all the possible reciprocal positions that the two duplexes could assume while interacting. From the bend angle we calculated the persistence length, that has to be compared to the one used in the coarse grained model.

From now on we calculated some mean quantities that are averaged over the positions of the floating duplexes that assure it to interact with the fixed duplex. This volume is calculated on the basis of the potential well just analysed. The volume $V$ that contains all such positions, is determined by the set of coordinates (x,y,z) that solve the inequality $(x, y, z)|E(x, y, z) \leq 0$. We then extracted the z coordinate as a function of the horizontal coordinates (x,y) that satisfy the inequality:

$$z = -D/F \tag{4.13}$$

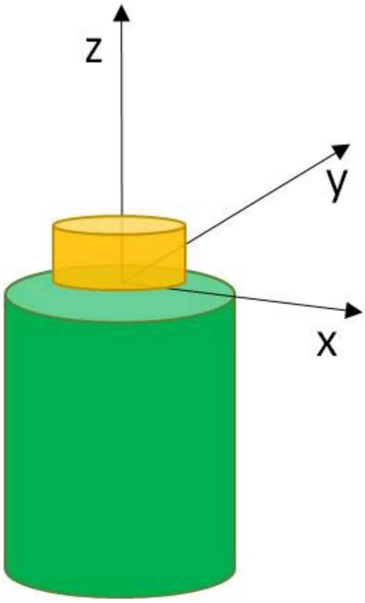


**Figure 4.22:** Surface in the (x,y,z) frame of reference that corresponds to $E = 0$.

In Fig.4.25 the surface with $E = 0$ is represented.

### 4.6.3 Calculation of the persistence length of linear aggregates of AllAT from the shape of the simulated stacking potential well

The average bend angle $\psi$ for every reciprocal position of the two duplexes was calculated as the mean over the angular range allowed for the thermal motion of one duplex if the position of the other was fixed. For simplicity we used a geometrical argument considering the duplexes as two rigid cylinders of diameter D and height L. The reference system used is the same as before, with the origin in the center of the circular face of the fixed rod. We wrote the distance of the center of the terminal face of the floating rod in cylindrical coordinates as a function of the height $z$ and the radial distance $r$ from the origin. First we projected the possible bend angles in a plane perpendicular to the terminal face of the fixed cylinder, and we calculated in this two-dimensional geometry the maximum angle between the principal axes of the cylinders when the floating cylinder is bent. Second, we recovered the tree-dimensionality of the system by averaging the bend angles over all the planes perpendicular to the terminal face of the fixed cylinder.

Because of the geometry of the system, we considered the bend on the left and on the right directions, as shown in Figs.4.23(a) and 4.23(b). From the bend on the left we know that the maximum allowed bend angle is

$$\psi_1 = arcsin\left(\frac{2z}{D}\right) \tag{4.14}$$

If we consider the bend on the right we obtain that the maximum bend angle is

$$\psi_2 = arctan\left(\frac{2z}{D-2r}\right) \tag{4.15}$$

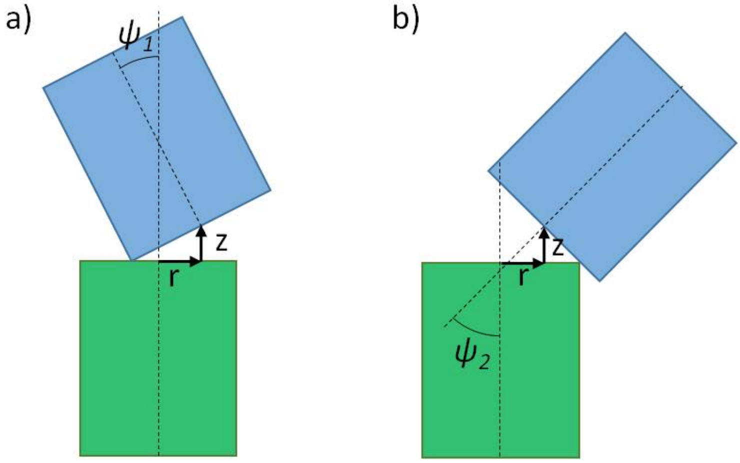


**Figure 4.23:** Maximum bend angle of a cylinder interacting with another identical cylinder at a random reciprocal distance individuated by the height $z$ and the radial distance $r$ in the plane of the terminal face of the fixed rod. Both bend directions (left and right) are considered.

The maximum bend angle depends on the bending direction. The directions for which the bend angle is $\psi_2$ can be individuated by an azimuthal angle $\psi^* = 2\pi - 2ArcCos\left(\frac{r^2-l^2+D^2/4}{Dr}\right)$ where $l^2 = \frac{D^2}{4} - h^2$. In Fig.4.24 the geometrical quantities considered are plotted.

$\psi_1$ and $\psi_2$ are the bigger bend angles as long as $h \leq h^* = \sqrt{Dr - r^2}$, because in the opposite case we just have to consider $\psi_1$ both for left and right bend. If we fix both the positions of the center of the terminal faces of the two interacting duplexes, we can calculate the average bend angle of one duplex if the other is kept fixed. This angle can be calculated as follows, if $h \leq h^*$:

$$\begin{aligned} \psi_{m1} &= \frac{\int_0^{\phi^*} d\phi \int_0^{\psi_1} \psi sin(\psi) d\psi + \int_{\phi^*}^{2\pi} d\phi \int_0^{\psi_2} \psi sin(\psi) d\psi}{\int_0^{\phi^*} d\phi \int_0^{\psi_1} sin(\psi) d\psi + \int_{\phi^*}^{2\pi} d\phi \int_0^{\psi_2} sin(\psi) d\psi} \\ &= \frac{\phi^*(sin(\psi_1) - \psi_1 cos(\psi_1)) + (2\pi - \phi^*)(sin(\psi_2) - \psi_2 cos(\psi_2))}{\phi^*(cos(\psi_1)) + (2\pi - \phi^*)(1 - cos(\psi_2))} \end{aligned} \tag{4.16}$$

If $h \geq h^*$:

$$\psi_{m2} = \frac{\int_0^{2\pi} d\phi \int_0^{\psi_1} \psi sin(\psi) d\psi}{2\pi \int_0^{\psi_1} sin(\psi) d\psi} = \frac{(sin(\psi_1) - \psi_1 cos(\psi_1))}{1 - cos(\psi_1)} \tag{4.17}$$

If we vary the position of the floating duplex inside the volume $V$ (the volume in which a duplex can move while interacting with the other fixed duplex), we find different average bend angles. The volume $V$ does not depend on temperature, while the probability of assuming one position inside $V$ changes as temperature is changed. As a consequence the average bend angle also changes. If we fix the temperature the floating duplex moves into the volume $V$ assuming the different positions with different probabilities depending on the value of the temperature and the interaction strength between

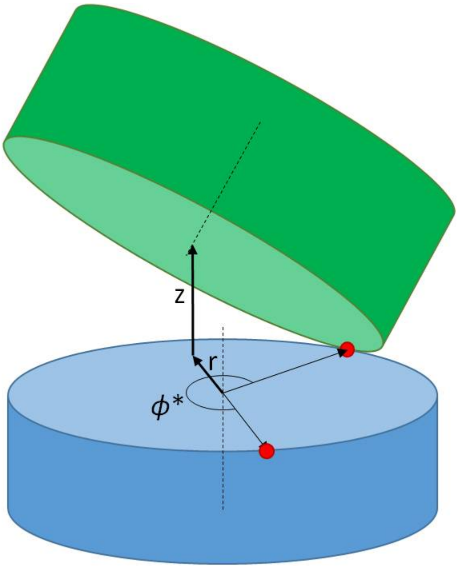


**Figure 4.24:** The azimuthal angle $\psi^*$ drawn on the circular end of the fixed cylinder is the angle for which the maximum bend of the floating duplex is $\psi_2$.

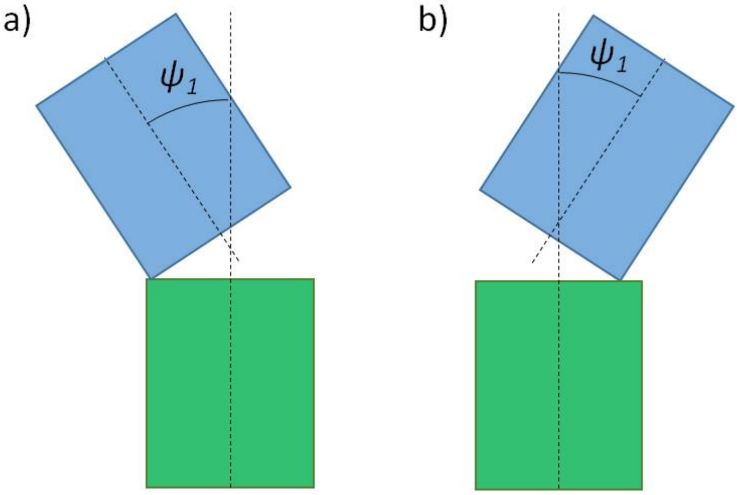


**Figure 4.25:** Scheme of the maximum bend angles when $h \geq h^* = \sqrt{Dr - r^2}$.

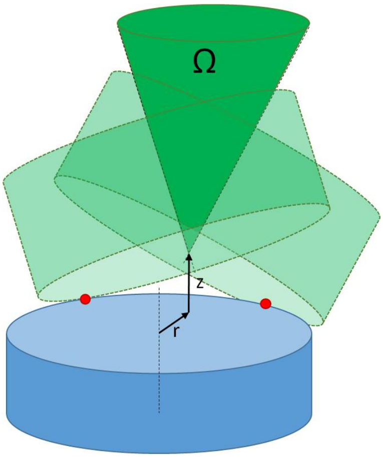


**Figure 4.26:** Angular volume accessible to a floating duplex interacting with a fixed duplex where the position of the center of the terminal face of the floating duplex is fixed.

the two duplexes. For every temperature we calculated the mean bend angle between the two duplexes averaging the bend angles $\psi_{m1}$ and $\psi_{m2}$ calculated for every position. These two angles are weighted by the angular volume accessible for the corresponding (z,r), that is the angular volume defined considering the azimuthal rotation of the floating duplex around the vertical direction parallel to the principal axis of the fixed cylinder but passing through the center of the terminal face of the floating cylinder. We thus proceed with computing such angular volume $\Omega$. The bigger is the bend angle, the bigger is the angular volume accessible for the cylinder. The angular volume allowed for the thermal fluctuation of an interacting duplex for $h \leq h^*$ is:

$$\begin{aligned}\Omega_1 &= \int_0^{\phi^*} d\phi \int_0^{\psi_1} sin(\psi) d\psi + \int_{\phi^*}^{2\pi} d\phi \int_0^{\psi_2} sin(\psi) d\psi = \\ &= \phi^*(1 - cos(\psi_1)) + (2\pi - \phi^*)(1 - cos(\psi_2))\end{aligned} \tag{4.18}$$

while for $h > h^*$ the angular volume is

$$\Omega_2 = \int_0^{2\pi} d\phi \int_0^{\psi_1} sin(\psi) d\psi = 2\pi(1 - cos(\psi_1)) \tag{4.19}$$

$\Omega$ is represented in Fig.4.26.

Then we should average the bend angle over all the possible positions of the duplex still interacting with the other one. The average must be weighted with the Boltzmann probability of finding the two duplexes in the considered configuration, that is equal to the probability of having such interaction energy that allows the duplexes to assume this configuration.

$$\psi_{bend} = \frac{\int_V e^{-\beta E(z,r)} \psi_{m1}(r, z, \phi) \Omega(r, z, \phi) d^3V}{\int_V e^{\beta E(z,r)} \Omega(r, z, \phi) d^3V} \tag{4.20}$$

The region of the integration is the same for both integrals. There are some constraints that have to be considered:

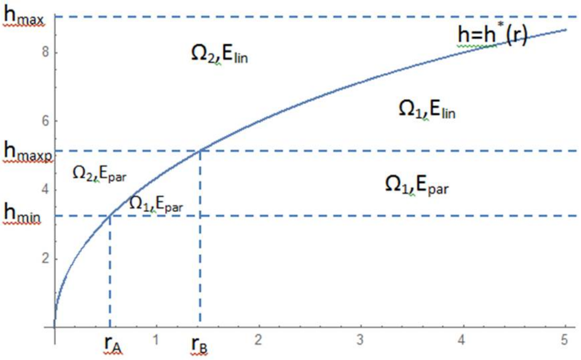


**Figure 4.27:** Constraints for the integration of equation 4.20. $E_{par}$ means that in those regions of integration the parabolic fit of equation 4.8 for the potential has to be used, while $E_{lin}$ stands for the linear fit.

- for $3,2\,\AA\,(= z_{min}) < z < 4,7\,\AA(= z_{maxp})$ the E(z) is fitted with a parabola.
- For $z_{maxp} < z < 8,1\AA(= z_{max})$ the E(z) is fitted with a linear function. $z_{max}$ is the end-to-end distance at which $E = 0$.
- The region $3,2 < z$ are not considered.
- When $z < z^*$ the angular volume $\Omega$ is calculated as $\Omega_1$, while for $z > z^*$ we have to use $\Omega_2$.

$r_A$ and $r_B$ are calculated from $h^* = h_{min}$ and $h^* = h_{maxP}$:

$$\begin{aligned} r_A &= \frac{D - \sqrt{D^2 - 4z_{min}^2}}{2} \\ r_B &= \frac{D - \sqrt{D^2 - 4z_{maxP}^2}}{2} \end{aligned} \tag{4.21}$$

$r_max = \sqrt{\frac{G}{H}}$ is the radial distance when the shear energy is null. In Fig.4.27 there is a sketch of the constraints that have to be considered in the integration.

In Fig.4.28 the average bend angle calculated from the model are shown. We noticed that if the depth of the potential well was kept fixed, the bend angles did not show any variation upon changing the temperature as happens for the angles extracted from the simulations (the one shown in table 4.3).

To calculate the persistence length of linear aggregates of duplexes we needed the mean bend angle between the axes of two consecutive duplexes. From the calculations shown above we found the bend angle between the axes of the circular faces of the terminal bases of the two cylinders. Because of the duplexes are not stiff as cylinders we needed to add the contribution due to the internal bend of the AllAT double helix. We extracted the internal bend angle of the helix as the angle between the axis of the terminal couple of bases and the axis of the central couple of bases. In table 4.4 we listed the values of the bend angle internal to the double helix structure at different temperatures.

A linear aggregate of N duplexes was then considered as a freely jointed chain made by 2N stiff monomers with fixed angles between them. The values of the angles were alternatively $\psi_{bend}$ and $\psi_{int}$. In Fig.4.29 this approximation is sketched.

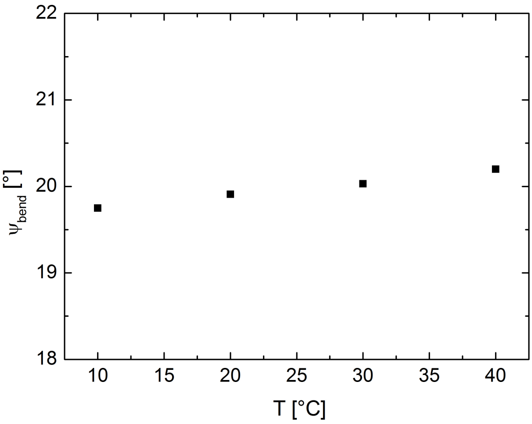


**Figure 4.28:** Average bend angles calculated from the 3D potential well

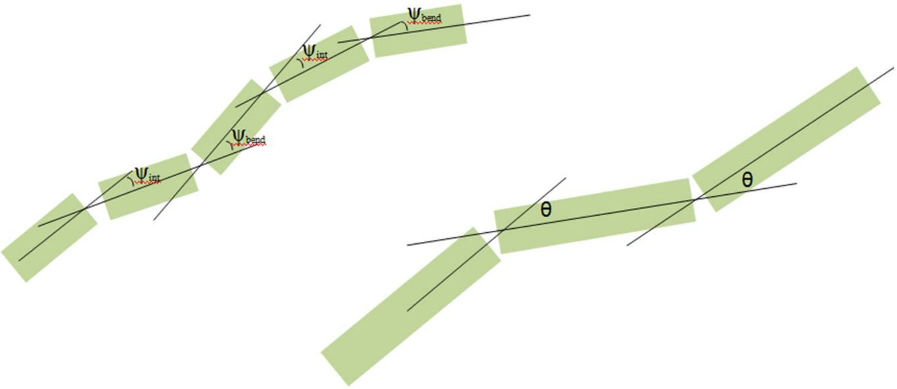


**Figure 4.29:** Sketch of two different ways of representing a chain of cylinders with the same persistence length. The monomers of the chain on the left are alternatively bonded with angles $\psi_{int}$ and $\psi_{bend}$. Instead the chain on the right is made on monomers of doubled length and that are all connected with the same angle, $\theta$.

| T [°] | $\psi_{int}$ [°] |
|---|---|
| 10 | 20 |
| 20 | 20,6 |
| 30 | 20,4 |

**Table 4.4:** Average bend angles internal to the double helix structure of AllAT extracted from atomistic simulations at different temperatures and averaged over the two system configurations.

The persistence length of the linear aggregates of cylinders was calculated from the definition for freely jointed chain. It is defined from the scalar product of the two vectors individuated by two monomers having a distance of $s - s'$ monomers:

$$< \mathbf{u}(s) \cdot \mathbf{u}(s') >= e^{-\frac{|s-s'|}{L_P}} \tag{4.22}$$

In the average bend angle between two monomers is $\gamma$, the scalar product is also equal to $< \mathbf{u}(s) \cdot \mathbf{u}(s') >= cos(\gamma)^{\frac{|s-s'|}{b}}$ where $b$ is the monomer length. In our system there are two characteristic bend angles, $\psi_{bend}$ and $\psi_{int}$, and the length of the monomer is then half of the length of the duplex. Within $s - s'$ monomers, half of the junctions have an angle $\psi_{bend}$ and half $\psi_{int}$. The scalar product becomes:

$$< \mathbf{u}(s) \cdot \mathbf{u}(s') >= cos(\psi_{bend})^{\frac{2|s-s'|}{2b}} cos(\psi_{int})^{\frac{2|s-s'|}{2b}} \tag{4.23}$$

The persistence length $L_P$ is then

$$L_P = -[ln(cos < \psi_{bend} > cos < \psi_{int} >)]^{-1} \tag{4.24}$$

To validate the results of the calculations we compared the results with the persistence length calculated directly from the simulations. In Fig.4.30 both quantities are shown at different temperatures. We noticed that the behaviour at different temperature is similar because in both cases the $L_P$ do not change upon varying the temperature. The values however are slightly different, of about the 10%. There are some aspects that can be verified, like the compatibility between the definition of the axes adopted in the atomistic simulations and in the coarse grained model. The persistence length calculated from the reconstruction of the potential is a little larger than the one extracted from the simulations, to indicate that probably the geometric models we built does not capture all the bend angles that two duplexes can assume. For example the presence of the minor and major grooves in the double helix yield bigger bend angles in specific directions than the angles found in the cylinder approximation.

### 4.6.4 Calculation of $w$ and $< E >$ from the atomistic potential well

In order to give an estimate of the strength of the stacking energy between blunt ended duplexes we compared the prediction of a coarse grained model of LC phases of duplexes with the atomistic simulations that describe the interactions between two duplexes. As a result, we calculated the persistence length and the stacking binding energy with both models. To fully compare the two models we need to calculate other two parameters from the atomistic simulations. The parameter are $w = 2\frac{v_b}{v_d}$ and the value that should be put into the coarse grained model for the depth of the square potential well. In fact we should approximate the soft potential well of the atomistic simulations

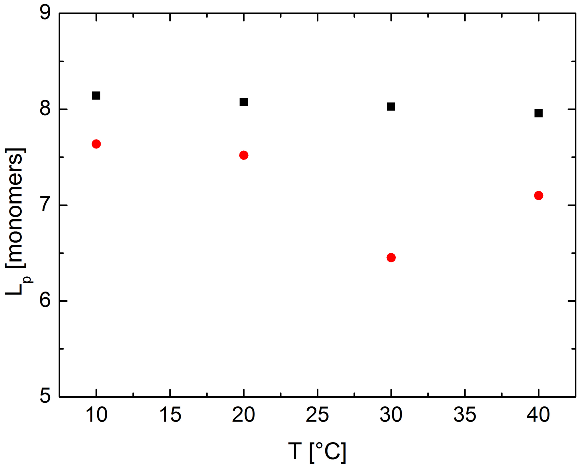


**Figure 4.30:** Comparison between the persistence length calculated from the reconstruction of the potential well (black squares) and the one extracted from the all atom simulations (red circles)

to an effective square well potential required by the coarse grained model. We chose to use a square well potential with depth the interaction energy weighted over the thermal energy and the number of conformations that the duplexes can assume. The number of conformations at fixed energy are proportional to the angular volume $\Omega$. The effective energy then becomes:

$$< E >= \frac{\int_V e^{-\beta E(z,r)} E(z,r,\phi)\Omega(z,r,\phi) d^3V}{\int_V e^{-\beta E(z,r)} \Omega(z,r,\phi) d^3V} \tag{4.25}$$

where the volume $V$ is the volume that the center of the terminal couples of bases of one duplex can explore while interacting with the terminal of one other duplex. $\Omega$ assumes the expression of $\Omega_1$ or $\Omega_2$ as the coordinates (z,r) vary within $V$.

The parameter $w$ is calculated as follows. The bonding volume $v_b$ is calculated as the product of the volume $V$ and the angular volume that the duplex can assume weighted with the corresponding Boltzmann probability. The volume $v_d$ is the volume of the single duplex multiplied for the angular volume that it can access when free from the interaction with other duplexes. $w$ becomes then:

$$w = 2\frac{V}{V_{duplex}} \frac{\int_V e^{-\beta E(z,r)} \Omega(z,r,\phi) d^3V}{4\pi \int_V e^{-\beta E(z,r)} d^3V} \tag{4.26}$$

### 4.6.5 Results

We varied the depth of the 3D potential well reconstructed from the atomistic simulations of AllAT duplexes to calculate the persistence length of the linear aggregates of duplexes, the effective free energy $< E >$ and the parameter $w$ that best reproduce the experimental concentration phase coexistence through the coarse grained model developed by De Michele et al. In table 4.5 we listed the best set of parameters found as a function of the temperature.

| T [°] | $d_E$ | $L_p$ [monomers] | $w$ | $< E >$ | $\beta\Delta G$ |
|---|---|---|---|---|---|
| 18 | 4,26 | 8,76 | 0,0068 | -2,87 | -0,012 |
| 25 | 3,91 | 8,34 | 0,0073 | -2,37 | -0,905 |
| 31 | 3,19 | 7,69 | 0,0080 | -1,69 | -2,022 |
| 37 | 2,84 | 7,29 | 0,0089 | -1,35 | -2,523 |

**Table 4.5:** The best set of parameters that can be used into the coarse grained model to fit the experimental data of the concentration phase coexistence of the isotropic-nematic phases of AllAT duplexes. From left to right the columns correspond to temperature, depth of the potential well, persistence length of the linear aggregates of duplexes, parameter $w$, proportional to the ratio between the bonding volume and monomer volume, and depth of the square well potential of the coarse grained model.

The experimental measurements cannot be not well reproduced using the parameters obtained from the all atom simulations, even if the depth of the potential well is reduced to 10% of its initial value. Even if we strongly decrease the depth of the potential well, the persistence length and the absolute value of the effective energy $< E >$ remains too high to yield to phase boundaries compatible to the experimental ones. The cause of this phenomenon comes from the fact that the potential well is too sharp, so even when the depth is decreased the probability of visiting a region of the well does not change much. As a consequence the average energy experience a small change and also the other parameters. It is possible that the sharpness of the potential well comes from the restraints used in the simulations from which we extracted the shape of the potential well: the Watson-Crick interactions between the complementary couples of bases were reinforced in order to keep all of the couples of bases paired for both duplexes during the simulations. This is an extremely non-physical behaviour because at the temperature sampled it has been noticed that the terminal couples of bases of duplexes in solution has a nonzero probability of opening. If those restraints were released, the terminal bases would be more free to interact with the terminal bases of the other duplex for a range of end to end and shear distances longer than the one obtained while the restraint is applied. Unfortunately it is extremely hard to perform simulations without such restraints because the timescale of the dynamics of the opening and closure of the bonds of the terminal couples of bases is much longer than the typical duration of the simulation done so far. We will try to perform such simulation in a dedicated computer.

Another possible reason for the incompatibility of the parameters calculated with the model and the simulations is that we approximated the monomers of the model as two cylinders. The monomers considered in the coarse grained model are, in fact, cylinders with smoothed angles. We tried to represent this shape with a decrease of the diameter o the cylinder. We tried to reduce the diameter of the duplex while calculating $L_p$, $< E >$ and $w$. Using such procedure we obtained the values of parameters closer to the ones of table 4.2. In table 4.6 we listed the binding free energy and the persistence length obtained for a diameter of $16\ \AA$ at different temperatures.

| T [°] | $d_E$ | $L_p$ [monomers] | $\beta\Delta G$ |
|---|---|---|---|
| 18 | 3,83 | 6,14 | 0,11 |
| 25 | 3,34 | 5,55 | -0,82 |
| 31 | 2,70 | 4,84 | -1,73 |
| 37 | 1,42 | 3,82 | -2,75 |

**Table 4.6:** Persistence length and binding energy calculated at different temperatures and amplitude of the potential for two interacting cylinders of diameter 16 $\AA$.

## 4.7 Conclusions

The aim of the work presented in this chapter is to give an estimate of the strength of the stacking interaction between two blunt ended duplexes of AllAT. For this purpose we studied the LC phases and in particular the nematic phase of two racemic mixtures of AllAT at two different ionic strengths. We measured the concentration at the boundaries of the isotropic-nematic phase coexistence and the order parameter at various temperatures (section 4.3.1).

In sections 4.5.1 and 4.5.2 we tried to fit the data with the expectations of a coarse grained model that predicts the phase boundaries and the order parameter of smoothed cylinders with a sticky interaction at the ends as a function of some parameters, that are the geometrical dimensions, the persistence length of linear aggregates of monomers and the binding energy between couples of monomers. As a result we found two unique sets of parameters that best reproduce the data at a given temperature of 20°C, one set for every ionic strength of the racemic mixture.

We then tried to fit the data at the different temperatures scanned (section 4.5.4). The sticky interaction is described by a square well potential, that provides the persistence length to be constant at all the temperatures. With this assumption we could not reproduce the observed behaviour of the phase boundaries as a function of the temperature. To achieve a good compatibility with the data, we were forced to change the persistence length as a function of the temperature. This assumption is supported by two reasons: the stacking interaction is a continuously decaying potential and its strength depends on the temperature. However we could not understand which one of these two phenomena caused the $L_p$ to change with the temperature, because the persistence length and the binding free energy were arbitrarily chosen. Thus we tried to reproduce both quantities through the all-atom simulations, to understand if a continuously decaying potential could yield the temperature dependence of the persistence length or if it was necessary to decrease the potential amplitude to achieve such behaviour of the persistence length.

The all-atom simulations first stated that a continuously decaying stacking potential brings to the formation of linear aggregates of duplexes whose persistence length is constant (section 4.6.1) in the range of temperatures scanned (10-$40^{circ}$). As it is known that the stacking potential used in such simulations is overestimated and because from the simulation emerged that it did not change between 10 and 40°, we developed some calculations on the basis of the stacking potential obtained from the simulations (section 4.6.3). Specifically, we varied the amplitude of the stacking potential as a function of the temperature and as a result we recovered the temperature-dependence of the persistence length.

At the same time we tried to reproduce the persistence length and the binding free energy that were used into the coarse grained model to best fit the data, by varying the amplitude of the stacking potential well and by fixing the geometrical dimensions of the duplex as the ones used in the model. Through this procedure, even if the amplitude was decreased to the 10% of its initial value, the expected values of the two parameters could not be obtained. A possible reason is that the bonds between the couples of bases were restrained in order to contain the simulations costs.

As the monomers of the coarse grained model are smoothed cylinders, whose diameter on the top is smaller than the diameter in the middle, we reduced the diameter of the duplexes in the calculations based on the all-atom stacking potential and we succeeded in recovering the $L_p$ and $\beta\Delta G$ that were obtained from the coarse grained model (section 4.6.5. Because of the diameter has to be largely reduced from the physical diameter of the duplex, as a further step we planned to run a long simulation without the constraints on the terminal couples of bases to see if the stacking potential would change. We will try to use this potential and the diameter obtained from the coarse grained model to verify if they could yield values of $L_p$ and $\beta\Delta G$ compatible with the one found from the model.

CHAPTER 5

# Matherials and methods

This chapter provides a description of the techniques used to characterize the nematic phase of DNA.

The melting temperature, concentration and birefringence of the nematic phases studied were measured through observations made by Polarized Transmission Optical Microscopy (PTOM) and Polarized Reflection Optical Microscopy (PROM). To observe LC phases of DNA it is necessary to use microscopy techniques, amongst which optical microscopy. This latter techniques produces a magnified image of the sample observed, provides a resolution of the details of the image and makes them visible through naked eye or with a camera. DNA molecules are too small to be directly observed through a microscope, but it is possible to obtain information about their disposition within the sample through the use of polarizers, that enhance the anisotropy of the molecules.

## 5.1 Measurement of the melting temperature

With a transmission optical microscope (Nikon, Optipho T2-pol) and a camera (Nikon Digital Sight Ds-U1) we collected images of the samples at different temperatures in a range between 10 to 45°C between crossed polarizers and in Koeler illumination regime. A sample with a nematic phase is bright if observed between crossed polarizers, while an isotropic phase is dark. If we observe a nematic phase through crossed polarizers while increasing the temperature, we see the bright region getting smaller as the nematic phase progressively melts. We defined the melting temperature of the sample as the temperature at which all the phase is melted. The temperature of the sample is controlled through a heater (Instec STC 200D) connected to the microscope.

## 5.2 LC optical anisotropy

The DNA double helix has an optical anisotropy because it has different polarizabilities along the main axes of the helix and in the perpendicular plane. A solution of double helices of DNA in the isotropic phase has no resulting optical anisotropy because the molecules have random orientations. In the nematic phase on the other hand the helices have all the same average orientation so the birefringence is the sum of the birefringencies of the helices. The birefringence of a uniaxial medium is defined as

$$\Delta n = n_{||} - n_{\perp} \tag{5.1}$$

where $n_{||}$ is the refraction index along the optical axis and $n_{\perp}$ is the one along the direction perpendicular to the optical axis. The sum of the two indexes lead to the index ellipsoids (see Fig.5.1).

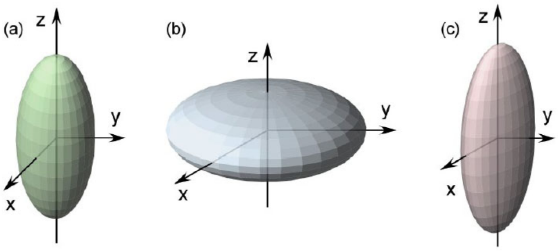


**Figure 5.1:** Ellipsoids that represent the value of the refractive index in a a) positive uniaxial medium, b) negative uniaxial and c) biaxial medium.

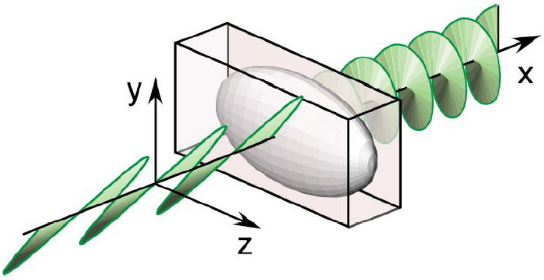


**Figure 5.2:** Linearly polarized light changes the polarization due to the passage through a birefringent sample.

From now on the monoaxial phases are described. They are media with two principal refractive indexes and just one optical axis, that is individuated by the direction perpendicular to plane whose intersection with the ellipsoid is a circle. In such medium the incoming light is split into two directions, the ordinary and the extraordinary directions. The two components propagate with different velocities because the refraction indexes ($n_o$ and $n_e$) are different, so after passing through a sample of thickness d there is a phase difference $\delta$ between the two components:

$$\delta = \frac{2\pi}{\lambda}(n_e - n_o)d \tag{5.2}$$

where $\lambda$ is the electromagnetic wavelength in the vacuum. If the optical axis of the medium and the direction of the incident light do not coincide, we can define the refractive indexes along the direction of the incident light and the perpendicular direction:

$$\begin{aligned} n_o &= n_\perp \\ n_e &= \frac{n_{||} n_\perp}{\sqrt{n_{||}^2 cos^2\phi + n_\perp^2 sin^2\phi}} \end{aligned} \tag{5.3}$$

where $\phi$ is the angle between the optical axis and the direction of the light. Typically, if the incident light that is linearly polarized is converted into elliptically polarized after passing through the birefringent sample. If the transmitted light is collected after a linear polarizer then the intensity is a function of $\phi$ and $\delta$.

$$I = I_0 sin^2(2\phi) sin^2(\delta/2) \tag{5.4}$$

The first factor modulates the intensity as the orientation of the sample changes, while the second is different for different wavelengths and is responsible for the color of the nematic phase. $I_0$ is the intensity recovered after the polarizer and $\phi$ is the azimuthal angle between the analyzer and the projection of the optical axis on the plane of the sample.

If a well ordered uniaxial nematic sample is studied with a microscope between two crossed polarizers, we can have different textures as a function of the orientation of the sample in the plane of observation. If the nematic director is perpendicular to the sample plane we have the omeotropic alignment, where the direction of the propagation of the light coincides with the optical axis. Here $\phi = 0$ and as a consequence $n_e = n_o$ and $\delta = 0$. The intensity transmitted through the sample in this configuration is $I = 0$ for all the azimuthal orientation of the sample. For a uniform planar alignment $\phi = 90°$ and $n_o = n_\perp$ and $n_e = n_{||}$, so $\Delta n = n_{||} - n_\perp$. The transmitted intensity varies as a function of $Sin^2(2\phi)$. It has a maximum for $\phi = 45°$ and a minimum for $\phi = 0$ and $\phi = 90°$. The sample is black when the optical axis is parallel to the direction of the polarizer or the analyzer, and the textures have a periodicity of $90°$ if the ample is rotated between two crossed polarizers

### 5.2.1 Concentration measurement

The solution of DNA between the two glass slides, of refractive index $n_1$, of thickness d has a mean refractive index (at a temperature at which the phase is isotropic) $n_2$. If a light beam goes onto such system we have a transmitted and reflected beam that depend on the angle of incidence.

$$T = \frac{1}{1 + Fsin^2(\delta/2)} R = 1 - T = \frac{Fsin^2(\delta/2)}{1 + Fsin^2(\delta/2)} \tag{5.5}$$

Where $F = \frac{4R_1}{(1-R_1)^2}$ is the contrast of the interference fringes, $R_1$ is the reflectivity of the first interface, between glass and DNA solution. The reflectivity depends on the two refractive indexes while the contrast depends only on the index refractive of the DNA solution:

$$\delta = \frac{4\pi n_2 dcos\theta}{\lambda} \tag{5.6}$$

where $\theta$ is the angle of incidence and $\lambda$ the wavelength of the light in vacuum. If the thickness d is known the refractive index $n_2$ can be measured from the reflected or transmitted light. In fact when $\delta = \pi \pm 2\pi m$ or $0 \pm 2\pi m$ with m integer then R and T are minima. We measured T and R for a wide range of visible wavelengths and we determined the product $n_2 d$. The wavelengths of the maxima (M) and minima (m) are:

$$\lambda_m = \frac{4n_2 dcos\theta}{2m} \lambda_M = \frac{4n_2 dcos\theta}{2m+1} \tag{5.7}$$

Because we do not know the m that correspond to the observed minima, we calculated the reciprocal of $\lambda$ both for maxima and minima of the reflected light, in such a way that the dependence from m disappeared.

$$\frac{1}{\lambda} = \frac{1}{\lambda_0} + \frac{1}{4n_2 dcos\theta} \tag{5.8}$$

From $\lambda$ and $\lambda_0$ we calculated the product $n_2 d$. The thickness d is then determined measuring the minima of T and R for the sealing oil, whose refractive index is known. The measurements are done with a reflection microscope and a spectrophotometer compatible with it. The magnification used was 50x with a theoretical aperture of 0,45 that corresponds to the collection of an angle of 27° in air. With this set-up the region of sample from which we collect the reflected light is big enough to average over many domains with different orientations. The thickness of the sample is determined by interpolating the thickness of two consecutive regions of oil. The crucial point of the measurement is the flatness of the cell: if the thickness is not uniform in the range of observation the contrast of the fringes is low and the calculation of the thickness of the sample has a great error. The variation of the oil refractive index as a function of the temperature is $dn/dT = -5 \cdot 10^{-4} K^{-1}$.

A crucial point of these kind of measurements is the relationship between the refractive index and the concentration of the solution of DNA. We assumed a linear relation between them:

$$n = n_{H_2O} + \frac{dn}{dc} c \tag{5.9}$$

It is possible to extract the refractive index from the concentration and vice versa. In literature the linear coefficient is in the range 0,168 to 0,185 ml/mg for concentrations up to 160 mg/ml, that are far below the concentrations of the nematic phase. To obtain a reliable linear coefficient some measurements of refractive index were done at known higher DNA concentrations and the average value obtained was 0,136 ml/mg.

It is difficult to estimate the error of a concentration measurements because it is very complex, and we can give an estimate of the errors that comes from the experimental techniques and the interpretation of the data. In summary, the error of the concentration measurements are high, about 5% of its value.

## 5.3 Measurements of transmission spectra

A spectrophotometer (Oceanoptics INC, SD 2000) was used to collect the spectra of the samples with the aim of measuring the concentration of DNA in the nematic phase. The spectrophotometer works in the visible range of the electromagnetic radiation and is connected to an inverted microscope.

## 5.4 Birefringence measurements

The birefringence of the nematic phase was measured with a Berek compensator (Olympus, U-CBT) combined with the microscope. The concentration of the phase was measured using an inverted microscope (Nikon Te200) through the Fabry-Perot interference.

# List of Figures

1.1 Sketch of a DNA nucleotide with the adenine as the nucleobase.[6] 1
1.2 Different kind of nucleobases found in DNA and RNA molecules. Uracyle substitutes thymine in RNA molecules. 2
1.3 Phosphodiesteric bond between two nucleotides. Nucleobases, that should be found on the right side of sugar, are not shown. 2
1.4 Sketch of a *double stranded*-DNA molecule 3
1.5 Different 3-D structures of double-stranded DNA and RNA. From left to right they are called *A-form*, *B-form* and *Z-form*. 3
1.6 Possible pairing interactions between the nucleobases. 4
1.7 Stacking and pairing forces between various kind of nucleobases at different temperatures [7]. 5
1.8 Consecutive nucleotides ideally put in the same bidimensional plane [8]. 6
1.9 Twist, slide e roll of a double helix of DNA [8] 6

2.1 Sequence of basis of M13mp18 ssDNA from 5′ to 3′. The starting point is arbitrary. 12
2.2 Number of possible contacts as a function of the number of bases involved in each contact for M13mp18 and for an average between 4 different random sequences with the same statistical distribution of bases as M13mp18. The error bars are the standard deviation of the number of contacts over the 4 random sequences considered. We neglected the contacts that involve less that 5 base pairs. 14
2.3 Contact map of M13mp18 (picture above) and for a random ssDNA sequence (picture below) with the same statistical distribution of bases as M13mp18. 15
2.4 Two adiacent monomers of a chain with a fixed angle between them. [14] 18
2.5 Comparison between hydrodynamic and gyration radii and the end to end distance of a polymer chain for different models. 23
2.6 Ratio between the hydrodynamic to the gyration radius of a linear chain of PNIPAM polymer from [15] at different temperatures 23
2.7 Scheme of a DLS apparatus. 27
2.8 Sketch of a single laser set-up for dual-focus experiments, equipped with: 1) laser head, 2) mirror, 3) adjustable zeroth-order halfwave plate, 4) polarizing cube, 5) beam displacer, 6) fiber coupler, 7a) long single-mode fiber, 7b) short single-mode fiber. [17] 29

2.9 Intensity correlation function for a sample of M13mp18 at 0,3 g/l dissolved in a solution of 50 mM NaCl at 10°C (red line) and 25°C (blue line). The curves are normalized. 31
2.10 Examples of fitted intensity correlation functions for a sample of M13mp18 at 0,3 g/l dissolved in a solution of 50 mM NaCl. The correlations were measured this chronological order at a) 5°C, b) 15°C, c) 25°C and d) 30°C. 32
2.11 Intensity correlation functions of poly-A at a DNA concentration of 0,3 g/l and 50 mM NaCl. Temperatures are indicated in the legend. 34
2.12 Normalized intensity correlation functions of poly-A at a DNA concentration of 0,3 g/l and 50 mM NaCl. Time is divided for the water viscosity at the corresponding temperature. 34
2.13 Hydrodynamic radius of M13mp18 at different temperatures for 10 mM NaCl buffer. The scattering angle was kept fixed at 90° for all measurements. 35
2.14 Hydrodynamic radius of M13mp18 at different temperatures for 50 mM NaCl. The measurements were done while decreasing the temperature. The scattering angle was kept fixed at 90° for all measurements. The magenta line is drawn to guide the interpretation of the data. 35
2.15 Hydrodynamic radius of M13mp18 at different temperatures for 50 mM NaCl. The measurements were done while increasing the temperature. The scattering angle was kept fixed at 90° for all measurements. 36
2.16 All the hydrodynamic radii measured with DLS of samples of M13 at different ionic strengths. Black squares refer to the sample at 10 mM of NaCl, red circles and green diamonds are measured from samples at 50 mM NaCl while decreasing the temperatures. The blue triangles come from the samples at 50mM NaCl measured with increasing temperature. 37
2.17 Average scattered intensity measured with DLS for a sample of 0.3 g/l at 50 mM NaCl at different temperatures. The error bars are the standard deviation of the intensity. 38
2.18 Hydrodynamic radius of M13mp18 at different temperatures measured with FCS at 50 mM NaCl while increasing the temperature. 39
2.19 Average hydrodynamic radius of M13mp18 at different temperatures measured with FCS at 50 mM NaCl while slowly decreasing the temperature. 39
2.20 Intensity correlation functions measured for the first sample at 15°C and 50 mM NaCl. 40
2.21 Melting temperature of the possible contacts of M13mp18 as a function of the length of they loop they would form. Contacts that involve a different number of bases are plotted in different colors. The color code is the same of 2.3 43
2.22 Sketch of the conformation of M13mp18 at 15°C after the closure of the two contacts ($\alpha$ and $\beta$, table 2.3) involving 11 base pairs. The most distant regions of the chain that close while decreasing the temperature from 15°C to 10°C are shown by crosses. Firstly the magenta contact closes, then the dark green, the light blue and then the light green. 44
2.23 Melting temperature of the possible contacts as a function of the length of the loop after the closure of contacts $\alpha$ and $\beta$. The color code is the same of 2.3 47

3.1 Uniaxial nematic phase where **a** is the axis of one molecule and **n** is the nematic director. 50

3.2 Cholesteric phase formed by helical molecules. The direction of the nematic vector is shown. 51

3.3 Smectic phase of kind A and C, characterized by two different average directions of the molecules. 51

3.4 Example of a hexagonal columnar phase. 52

3.5 Phase diagram of self aggregating elongated molecules. $\phi$ is the volume fraction of the molecules, $v_p$ the volume of the single molecule and $T$ the temperature of the system. Shadowed areas indicate the coexistence region, while the vertical dotted lines are the steric limit for the packing. Points from $a$ to $f$ stand for different effective temperatures of the sample.[23] 54

3.6 (a) Scheme of the cholesteric phase of a solution of lDNA. (b) Projection of the nematic director along the direction perpendicular to the helix axis, (c) section along the cleaving plane and (d) representation of the duplexes of the superior layer. [24] 55

3.7 Example of cholesteric texture of a solution of lDNA when the cholesteric axis is parallel to the plane of the sample (on the left) or perpendicular (on the right). [24] 56

3.8 Summary of the fundamental properties of the cholesteric phase observed in solutions of different short duplexes of DNA. For each are listed the name of the sequence, the number of bases, the structure of the terminals, the concentration of the nematic phase, the cholesteric pitch, its behaviour with temperature and concentration and the chirality of the phase. INC stands for increasing, while DEC for decreasing. L determines a left handed phase, and R a right handed phase.[26] 57

3.9 Example of one BE duplex (on the top) and one SE duplex (on the bottom) 58

3.10 Average length of the aggregates in monomer units as a function of $\phi$ for sequences of about 12 couples of bases and for different values of the interaction energy between the duplex ends. Symbols come from MC simulations, while the dotted lines from theoretical models. [28] 59

3.11 Steric interactions between two duplexes. [27] 59

3.12 Scheme of a duplex of DNA as a cylinder with helical lines of negative electric charges and positive counterions on the side. 60

4.1 Circular dichroism spectra of a solution of (a) right-handed AllAT (dotted black line), left-handed AllAT (solid black line), (b) a racemic mixture of the two enantiomers (red line) and the baseline (black line). All the solutions were concentrated at 0,2 g/l DNA. 64

4.2 Polarized optical microscopy image of a nematic phase of a) a solution of D-allAT at a concentration of 595 g/l of DNA with 1,8M of NaCl and b) racemic allAT with 1,8M of NaCl at a concentration of 580 g/l. The images were taken with crossed analy¡er and polarizer, so the black regions are the one with isotropic phase. c) Racemic allAT without added salt at a concentration of 590 g/l d) the same racemic sample rotated by 45°. All images were taken with crossed analyzer and polarizer. 65

4.3 $T_m$ vs DNA concentration for the racemic mixture of AllAT without added salt. The red straight line is the fitting line of the green dots and provide the phase boundary of the isotropic phase. The blue line is the phase boundary of the nematic phase calculated through the data plotted in Fig.4.4 as explained in the text. In between the blue and the red lines we observed isotropic-nematic phase coexistence. 67
4.4 $\Delta T$ vs $T_m$ for the racemic mixture of AllAT without added salt. The red line is the best linear fit of the data. 68
4.5 Plot of the melting temperature of the nematic phase as a function of DNA concentration for racemic (R) allAT at 1,8M of NaCl (light blue squares), D-allAT at 1,8M of NaCl (purple triangles) and L-allAT at 1,8M NaCl (blue triangles). The straight lines are obtained fitting the filled dots for every sequence. 68
4.6 $\Delta T$ vs $T_m$ for D-AllAT (purple triangles), L-AllAT (blue squares) and racemic R-AllAT (light blue squares). All the solutions are at 1,8 M NaCl. 69
4.7 Measured coexistence regions for racemic (R) allAT at X M of NaCl (green region), racemic (R) allAT at 1,8M of NaCl (light blue region), D-allAT at 1,8M of NaCl (purple region) and L-allAt at 1,8M NaCl (blue region). The vertical lines stand for the $\Delta T$ range of samples whose concentration has been shifted to the fitted one 70
4.8 * 71
4.9 Measured birefringence of the nematic phase at different temperatures and concentrations of racemic allAT without added salt (b) and c), full symbols). The intensity of the transmitted light between crossed polarizers is shown for the two samples (empty symbols, a), b) and c) ). The black dashed vertical line separates the nematic (on the left of the line) and the isotropic-nematic coexistence region (on the right side of the line). In the isotropic-nematic coexistence region we could not measure the birefringence because the nematic regions were too small, while the intensity was measured in a region that contained both isotropic and nematic domains. As the red dotted line in a) shows, the intensity in the coexistence region decreases linearly with the temperature, as it is expected for a quantity that depends on the volume fraction of the nematic phase. 72
4.10 Order parameter calculated from the birefringence of racemic allAT with 1,8M of NaCl (a),full symbols) and without added salt (b), full symbols). 73
4.11 Measured (full symbols) and calculated (empty symbols) order parameters as a function of temperature for the racemic allAT with 1,8M of NaCl a) and without added salt b). The dotted line indicates the temperature at which the melting of the phase starts. 76
4.12 Comparison between the concentration phase boundaries measured and calculated using the coarse grained model. The straight lines are the measured concentrations of the isotropic-nematic phase boundaries. The empty dots are the phase boundaries calculated from the model using a fixed persistence length for every value of binding energy, while the full dots are calculated with the same values of the parameters but with a variable persistence length. On top the data for R-AllAT are shown, while on bottom there are the data for R-AllAT with 1,8M NaCl. 79
4.13 $\beta\Delta G$ as a function of the inverse of the temperature for R-AllAT and R-AllAT with 1,8M NaCl. 80
4.14 PMF vs azimuthal angle for D-D and D-L configurations, at 20°. 82

4.15 Bend angle $\theta$ calculated from the atomistic simulations at 20°C for two D-D and L-D duplexes as a function of the azimuthal angle $\phi$. From red to blue the PMF decreases and the probability increases. The probabilities of both D-D and L-D configurations are normalized such that the sum of the probabilities over both the different bend and the azimuthal angles are equal to 1. 84
4.16 Persistence length $L_p$ averaged over D-D and L-D configurations at different temperatures. It is expressed in monomer units. 85
4.17 E as a function of the end to end distance between the two duplexes. $E_{DD}$ and $E_{LD}$ are the potentials of the interaction between respectively D-D and L-D duplexes at $\phi \simeq 180°$ for D-D and $\phi \simeq 260°$ for L-D. $E_{10DD}$ and $E_{10LD}$ are the potentials whose minima are normalized to the values at 10° averaged over all the azimuthal angle, obtained from the data plotted din Fig.4.14. 87
4.18 Average smoothed energy E vs end to end distance between two duplexes. The fitting function is a parabola for $3,2\,\mathring{A} \leq z < 4,7\,\mathring{A}$ and a straight line for $4,7\,\mathring{A} \leq z < 8,1\,\mathring{A}$ 88
4.19 Normalized $E$ vs x and y shear directions at 10°C for D-D and L-D configurations, at 10°. 89
4.20 Normalized average interaction potential vs the shear direction at 10°C. The mean is calculated over the configurations and the sampled shear directions. 89
4.21 Scheme of the frame of reference adopted for the reconstruction of the 3D potential well. 90
4.22 Surface in the (x,y,z) frame of reference that corresponds to $E = 0$. 91
4.23 Maximum bend angle of a cylinder interacting with another identical cylinder at a random reciprocal distance individuated by the height $z$ and the radial distance $r$ in the plane of the terminal face of the fixed rod. Both bend directions (left and right) are considered. 92
4.24 The azimuthal angle $\psi^*$ drawn on the circular end of the fixed cylinder is the angle for which the maximum bend of the floating duplex is $\psi_2$. 93
4.25 Scheme of the maximum bend angles when $h \geq h^* = \sqrt{Dr - r^2}$. 93
4.26 Angular volume accessible to a floating duplex interacting with a fixed duplex where the position of the center of the terminal face of the floating duplex is fixed. 94
4.27 Constraints for the integration of equation 4.20. $E_{par}$ means that in those regions of integration the parabolic fit of equation 4.8 for the potential has to be used, while $E_{lin}$ stands for the linear fit. 95
4.28 Average bend angles calculated from the 3D potential well 96
4.29 Sketch of two different ways of representing a chain of cylinders with the same persistence length. The monomers of the chain on the left are alternatively bonded with angles $\psi_{int}$ and $\psi_{bend}$. Instead the chain on the right is made on monomers of doubled length and that are all connected with the same angle, $\theta$. 96
4.30 Comparison between the persistence length calculated from the reconstruction of the potential well (black squares) and the one extracted from the all atom simulations (red circles) 98

5.1 Ellipsoids that represent the value of the refractive index in a a) positive uniaxial medium, b) negative uniaxial and c) biaxial medium. 104

5.2 Linearly polarized light changes the polarization due to the passage through a birefringent sample. 104

# List of Tables

2.1 Fitting parameters of the intensity correlations plotted in Fig.2.10 measured with DLS as a function of temperature and salt concentration. $R_H$ is the hydrodynamic radius, $\tau$ is the average decay time, calculated as in 2.5.5, and $\alpha$ is the stretching exponent. The measurements were done at a fixed scattering angle of 90° while increasing the temperature. 33

2.2 Enthalpy and entropy variation [22] for the closure of the intra-chain contacts of 12 and 11 couples of bases and loop of three non interacting bases of M13mp18. Only one half of the two complementary single stranded sequences is shown as contact sequence. The ionic strength at which the parameters have been calculated is 50 mM NaCl. 43

2.3 The two contacts with the highest melting temperature that involve regions of the chains that are more than 2000 bases far away along the chain. The reported base sequence is the sequence of one of the two complementary regions forming the contact. The other is complementary. The distance is the number of monomers (nucleotides) that form the loop between the two contact regions. $T_m$ is calculated as in 2.56. 44

4.1 Best sets of parameters that fit the experimental data (the I-N phase boundaries at 20°C and the order parameter) for the two racemic samples. On the right there are the concentration boundaries obtained with the listed parameters. 77

4.2 Best set of parameters used to reproduce the experimental measurements with the coarse grained model. The persistence length is variable. 80

4.3 Average bend angles and persistence lengths calculated from atomistic simulations at different temperatures and for the two D-D and L-D configurations. 85

4.4 Average bend angles internal to the double helix structure of AllAT extracted from atomistic simulations at different temperatures and averaged over the two system configurations. 97

4.5 The best set of parameters that can be used into the coarse grained model to fit the experimental data of the concentration phase coexistence of the isotropic-nematic phases of AllAT duplexes. From left to right the columns correspond to temperature, depth of the potential well, persistence length of the linear aggregates of duplexes, parameter $w$, proportional to the ratio between the bonding volume and monomer volume, and depth of the square well potential of the coarse grained model. 99

4.6 Persistence length and binding energy calculated at different temperatures and amplitude of the potential for two interacting cylinders of diameter $16\,\AA$. 100

# Bibliography

[1] T. Omabegho, R. Sha, and N. Seeman. A bipedal dna brownian motor with coordinated legs. *Science*, 324, 2009.

[2] H. Gu, J. Chao, S. Xiao, and N. Seeman. A proximity-based programmable dna nanoscale assembly line. *Nature*, 465, 2010.

[3] P. W. K. Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440, 2006.

[4] M T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *P. Natl. Acad. Sci. USA*, 103(16):6190–6195, 2005.

[5] T. P. Fraccia, G. Smith, G. Zanchetta, E. paraboschi, Y. Youngwooo, D. M. Walba, G. Dieci, N. A. Clark, and T. Bellini. Abiotic ligation of dna oligomers templated by their liquid crystal ordering. *Nat. Comm.*, 6(6424):1–7, 2015.

[6] J. D. Watson and F. H. C. Crick. Molecular structure of nuclei acids - a structure for deoxyribose nucleic acid. *Nature*, 171, 1953.

[7] P. Yakovhuk, E. Protozanova, and M. D. F. Kamenetskii. Base-stacking and base-pairing contributions into thermal stability of the dna double helix. *Nucleic Acids Res.*, 34(2):564–574, 2006.

[8] C. R. Calladine, H. R. Drew, B. F. Luisi, and A. A. Travers. *Understanding DNA - The molecule and how it works, 3rd edition*. Elsevier Academic Press, 2004.

[9] S. M. Bhattacharjee, A. Giacometti, and A. Maritan. Flory theory for polymers. *J. Phys. Condens. Matter.*, 25, 2013.

[10] T. Škrbić, A. Badasyan, T. H. Hoang, R. Pdgornik, and A. Giacometti. From polymers to proteins: the effect of side chains and broken symmetry on the formation of secondary structures within a wang–landau approach. *Soft Matter*, 2016.

[11] M. P. Taylor, W. Paul, and K. Binder. Phase transitions of a single polymer chain: A wang–landau simulation study. *J. Chem. Phys.*, 131, 2009.

[12] D. R. Tree, A. Muralidhar, P. S. Doyle, and K. D. Dorfman. Is dna a good model polymer? *Macromolecules*, 46, 2013.

[13] F. Latinwo and C. Schroeder. Model systems for single molecule polymer dynamics. *Soft Matter*, 7, 2011.

[14] I. Teraoka. *Polymer solutions - An introduction to physical properties*. John Wiley and Sons, 2002.

[15] C. Wu and S. Zhou. First observation of the molten globule state of a single homopolymer chain. *Phys. Rev. Lett.*, 77(14):3053–3055, 1996.

[16] M. Rubinstein. *Polymer Physics*. Oxford University Press, 2003.

[17] C. B. Muller, A. Loman, V. Pacheco, F. Koberling, D. Willbold, W. Richtering, and J. Enderlein. Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. *EPL*, 83, 2008.

[18] T. Dertinger, V. Pacheco, I. Von der Hocht, Hartmann, and J. Enderlein. Two-focus fluorescence correlation spectroscopy: A new tool for accurate and absolute diffusion measurements. *Chem. Phys. Chem.*, 8, 2007.

[19] J. SantaLucia Jr. and D. Hicks. The thermodynamics of dna structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33, 2004.

[20] H. X. Zhou. Loops in proteins can be modeled as worm-like chains. *J. Phys. Chem.*, 105, 2001.

[21] H. Jacobson and W. H. Stockmayer. Intramolecular reaction in polycondensations. i. the theory of linear systems. *J. Chem. Phys.*, 18(12):1600–1606, 1950.

[22] J. SantaLucia Jr. A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *P. Natl. Acad. Sci. USA*, 95, 1998.

[23] M. P. Taylor and J. Herzfeld. Shape anisotropy and ordered phases in reversibly assembling lyotropic systems. *Phys. Rev. A*, 43, 1991.

[24] F. Livolant and A. Leforestier. Condensed phases of dna: structures and phase transitions. *Prog. Polym. Sci.*, 21, 1996.

[25] M. Nakata, G. Zanchetta, B. D. Chapman, C. D. Jones, J.O. Cross, R. Pindak, T. Bellini, and N. Clarck. End-to-end stacking and liquid crystal condensation of 6-to-20-base pair dna duplexes. *Science*, 318(13):1276–1279, 2007.

[26] G. Zanchetta, F. Giavazzi, M. Nakata, M. Buscaglia, R. Cerbino, N. A. Clarck, and T. Bellini. Right-handed double-helix ultrashort dna yields chiral nematic phases with both right- and left-handed director twist. *P. Natl. Acad. Sci. USA*, 107(41):17497–17502, 2010.

[27] E. Frezza, F. Tombolato, and A. Ferrarini. Right- and left-handed liquid crystal assemblies of oligonucleotides: phase chirality as a reporter of a change in non-chiral interactions? *Soft Matter*, 2011.

[28] C. De Michele, T. Bellini, and F. Sciortino. Self-assembly of bifunctional patchy particles with anisotropic shape into polymers chain: theory, simulation and experiments. *Macromolecules*, 5(13):2563–2570, 2011.

[29] C. Maffeo, B. Luan, and A. Aksimentiev. End-to-end attraction of duplex dna. *Nucleic Acids Res.*, 2012.

[30] J. Sponer, J. Leszczynski, and P. Hobza. Hydrogen bonding and stacking of dna bases: A review of quantum-chemical ab initio studies. *J. Biomol. Struct. Dyn.*, 14(1):117–135, 1996.

[31] T. Kuriabova, M. D. Betterton, and M. A. Glaser. Linear aggregation and liquid-crystalline order: comparison of monte carlo simulation and analytic theory. *J. Mater. Chem.*, 20, 2010.

[32] L. Campos, N. Valls, L. Urpi, C. Gouyette, T. Sanmartin, M. Richter, E. Alechaga, Santaolalla. A., R. Baldini, M. Creixell, R. Ciurans, P. Skokan, J. Pous, and J. A. Subirana. Overview of the structure of all-at oligonucleotides: Organization in helices and packing interactions. *Biophys. J.*, 91(3):892–903, 2006.

[33] N. Katsonis, E. Lacaze, and A. Ferrarini. Controlling chirality with helix inversion in cholesteric liquid crystals. *J. Mater. Chem.*, 22, 2012.

[34] F. Tombolato and A. Ferrarini. From the double-stranded helix to the chiral nematic phase of b-dna: A molecular model. *J. Chem. Phys.*, 122:054908, 2005.

[35] S. Pieraccini, S. amd Masiero, A. Ferrarini, and G. P. Spada. Chirality transfer across length-scales in nematic liquid crystals: fundamentals and applications. *Chem. Soc. Rev.*, 40:258–271, 2011.

[36] M. Vallazza, M. Perbandt, S. Klussman, W. Rypniewski, H. M. Einspar, V. A. Erdmann, and Ch. Betzel. First look at rna in l-configuration. *Acta Cryst.*, D60:1–7, 2004.

[37] M. Rossi, G. Zanchetta, S. Klussmann, N. A. Clarck, and T. Bellini. Propagation of chirality in mixtures of natural and enantiomeric dna oligomers. *Phys. Rev. Lett.*, 110, 2013.

[38] H. Schneider T. Nazarenko V. Nastishin, Y. A. Liu. Optical characterization of the nematic lyotropic chromonic liquid crystals: Light absorption, birefringence, and scalar order parameter. *Phys. Rev. E*, 72, 2005.

[39] S. Zhou, Y. A. Nastishin, M. M. Omelchenko, L. Tortora, V. G. Nazarenko, O. P. Boiko, T. Ostapenko, T. Hu, C. C. Almasan, S. N. Sprunt, J. T. Gleeson, and O. D. Lavrentovich. Elasticity of lyotropic chromonic liquid crystals probed by director reorientation in a magnetic field. *Phys. Rev. Lett.*, 109, 2012.

[40] V. R. Horowitz, L. A. Janowitz, A. L. Modic, P. A. Heiney, and P. J. Collings. Aggregation behavior and chromonic liquid crystal properties of an anionic monoazo dye. *Phys. Rev. E*, 72, 2005.

[41] A. Kumar. Calculation of optical parameters of liquid crystals. *Acta Phys. Pol. A*, 112(6):1213–1221, 2007.

[42] S. Chandrasekhar. *Liquid crystals*. Cambridge University Press, 1993.

[43] A. V. Kityk and P. Huber. Thermotropic nematic and smectic order in silica glass nanochannels. *Appl. Phys. Lett.*, 97, 2010.

[44] M. R. Puica. Lyotropic liquid crystals at phase transitions. *Romanian Reports in Physics*, 58(4):491–518, 2006.