UNIVERSITÀ DEGLI STUDI
DI MILANO

Doctoral Thesis

Selective Assembly, Phase Transitions and Molecular Kinetics of DNA oligomers

---

Settore Scientifico Disciplinare FIS/07

Author:
Simone Di Leo

Supervisor:
Prof. Tommaso Bellini

Co-Supervisor:
Dr. Tommaso P. Fraccia

To my women
Nicol and Teresa

"*Partendo dalla capacità di ridurre i fenomeni fisici all'essenziale, gli uomini hanno sviluppato la fisica degli ultimi secoli. E la fisica è diventata così potente e ricca da poter nuovamente introdurre nei propri modelli la complessità e il disordine, ciò che Galileo era stato costretto ad escludere.*"

Giorgio Parisi

UNIVERSITÀ DEGLI STUDI DI MILANO

Abstract

Facoltà di Scienze e Tecnologie
Dipartimento di Fisica "Aldo Pontremoli"

Doctor of Philosophy

Selective Assembly, Phase Transitions and Molecular Kinetics of DNA oligomers

by Simone Di Leo

In the last few years it has been shown that the spontaneous self-assembly process of short DNA and RNA duplexes into liquid crystal ordering is a likely potential route that led to the formation of first nucleic acids able to support biological activities. In particular, it has been experimentally demonstrated that liquid crystal domains behave as suitable micro-reactors to trigger polymerization between the stacked and not initially chemically linked short nucleic acids. Even paired mononucleotides at high enough concentration exhibit liquid crystal ordering, unveiling the crucial role of Watson-Crick selectivity and stacking attractive interactions among base pairs. In such a possible prebiotic context, DNA sequences with both random nucleobases sequence and length are likely to be formed. Surprisingly, it has been shown that even random DNA sequence of fixed length can support liquid crystal ordering at high concentration.
The aim of this PhD thesis is to extend the knowledge of DNA liquid crystals self-assembly in the following four directions.
First, I explored the selectivity of interaction in nucleic acids solutions of random-sequence DNA oligomers of different length L. The combination of experimental results and a suitable developed theoretical model revealed a not negligible percentage of perfect duplexes.
Second, I investigated the process that leads to the onset of the nematic liquid crystal phase in aqueous solutions of DNA duplexes. The combination of static light scattering experiments and computer simulations made possible the study of both aggregation and local ordering of DNA duplexes in the isotropic phase, where no positional order is developed, and in proximity of the isotropic-nematic phase boundary. This study gives an insight of the role on the development of local orientational order among DNA duplexes both far and in proximity of the isotropic-nematic phase boundary.
Third, I studied the diffusion of short DNA duplexes with attractive and repulsive interactions in the isotropic phase as a function of temperature. I found that the temperature dependence of diffusion coefficients reflects via an Arrhenius law the interduplex attractive interactions, whereas diffusion of repulsive duplexes is partially well described in terms of repulsive hard spheres.
Fourth, I investigated phase diagrams of mixtures of DNA single strands and duplexes with various polycations that show liquid-liquid phase separations. This phenomena leads to the onset of a concentrated but still liquid phase of polyelectrolytes, called coacervate, in a bulk phase where polyelectrolytes are diluted. The most surprising result I found, it is the insurgence of liquid crystals in coacervates with 12 nucleobases long random DNA oligomers and polylysine at different ionic strengths.
I believe that this PhD thesis adds important pieces to the self-assembly of nucleic acids puzzle, and in particular it shows how randomness of nucleic

acids is not an impasse to both hybridization of defectless duplexes and liquid crystal ordering.

The Chapters in the thesis are organized as follows:
Chapter 1: a brief introduction to DNA structure, fundamental interactions and self-assembly of liquid crystals.
Chapter 2: overview of light scattering and fluorescence recovery after photobleaching experimental techniques.
Chapter 3: contains melting UV experiments and theoretical model of random DNA oligomers solution.
Chapter 4: reports the static light scattering experiments and computer simulation results that allow to measure the aggregation of short DNA duplexes.
Chapter 5: reports the dynamic light scattering experiments on solution of repulsive and attractive short DNA duplexes.
Chapter 6: reports phase diagrams of mixtures of DNA oligomers with polycations that show liquid-liquid phase separation.

The work described in **Chapter 3** is published in:
Di Leo S., Marni S., Plata CA, Fraccia TP, Smith GP, Maritan A., Suweis S. and Bellini T. (2022). Pairing statistics and melting of random DNA oligomers: Finding your partner in superdiverse environments. PLoS Comput Biol 18(4): e1010051. https://doi.org/10.1371/journal.pcbi.1010051

Cover illustration:
Liquid crystal ordering in aqueous solution of random DNA with L=12 and polylysine.

Contents

Chapter 1

DNA and Self Assembly into Liquid Crystals

The aim of this chapter consists on introducing the main character of this story, the desossiribonucleic acid (DNA). In this Chapter, the structure and the main features of DNA along with a a brief description of Liquid Crystals (LC) are introduced.

1.1 DNA helix structure

The DNA macromolecule is the most important component of the smallest live unit, the cell. Indeed, it contains all the genetic information that are necessary to establish and regulate cell life. Here, the main interest consists on describing its structural properties such as double helix configuration, base paring and base stacking.
DNA is a weakly heterogeneous polymer formed by a chain of nucleotides, which are composed by a sugar, a phosphate group and a nitrogenous base. All the information that the DNA contains are given by sequence of 4 different nitrogenous bases: thymine (T), adenine (A), cytosine (C) and guanine (G). Polynucleotides are hence formed by a sequence of nucleotides which are defined by the nitrogenous bases sequence. A specific DNA sequence is conventionally written with the direction from $5'$-end to $3'$-end, where the number 5 and 3 refer to the carbon atom position of the sugar where the phosphate group and the hydroxyl group are linked, respectively. The DNA polynucleotides forming the DNA duplex are antiparallel, i.e. $5'$ terminal of a DNA sequence is linked to the $3'$ terminal of the other one, as shown in Fig.1.1. The shape of the DNA duplex is determined by several constraints: hydrophobicity and thickness of nitrogenous bases; local rigidity of the sugar-phosphate chain; narrow range distance between adjacent phosphates or sugars. The only shape that satisfies all such topological and physical proprieties is the helix[1]. Indeed, in such configuration the hydrophobic nitrogenous bases are well screened by outside water molecules. Three different kind of helices that differ for chirality and number of phosphates per turn are allowed. In this work it has been used the most common helix form, i.e. the B-form. B-form helix is right-handed and it has 10 phosphate per turn, corresponding to a length of ≈ 3.4nm, whereas the diameter of DNA is ≈ 2nm.

1.1.1 Fundamental interactions

The stability of DNA double helix is given by a combination of intermolecular forces between the two DNA sequences: stacking forces and the base pairing forces. Base pairing forces are due to the formation of hydrogen bonds between the nitrogenous bases. There are various rules that describe what atoms of two nitrogenous bases are involved in H-bonds formation, like for example Hoogsteen and Watson-Crick (WC) rules[1]. The most energetic

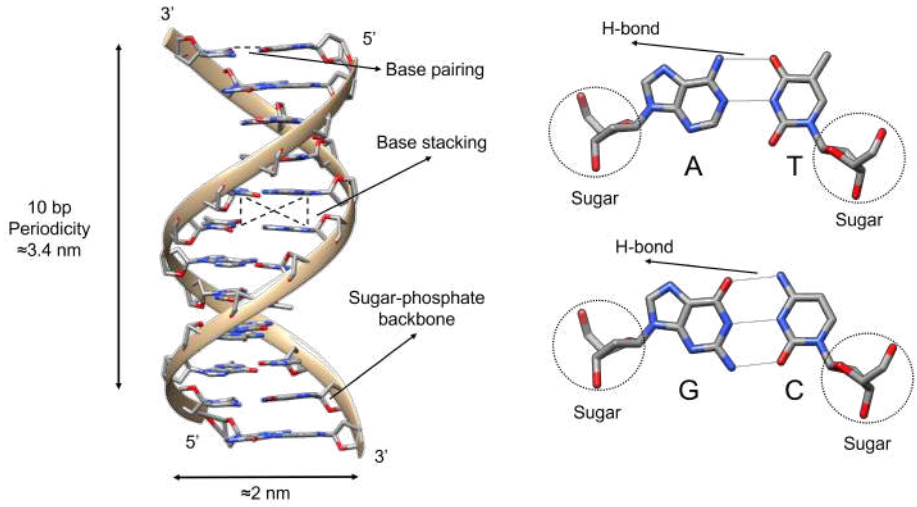

Figure 1.1: Left Panel: the double helix structure of B-DNA, with fundamental interactions and dimensions. Right panel: chemical structure of the four DNA nitrogenous bases along with sugar, showing the Watson-Crick pairing.

pairing is described by WC rules, where the pairing between the nitrogenous bases A-T and C-G occur with 2 and 3 H-bonds, respectively, as shown in Fig.1.1.

Base-stacking forces occur between adjacent base pairs and provide the main contribution to double helix stability. Stacking energies are given by a combination of different contributions: hydrophobicity of nitrogenous bases, that tend to maximize the superposition of interaction surface; electrostatic forces due negative charge surface (electrons delocalized on aromatic rings) that tend to minimize the interaction surface; attractive dipole-induced dipole and induced dipole-induced dipole interactions[2].
The total stacking energy between two consecutive base-pairs in a double helix can vary from -6 t o -17kcal/mol depending on the nucleotides pairs involved[3, 4, 5, 6]. Stacking energies play a crucial contribution in aggregation leading to the onset chromonic-type DNA structures[7] and LC phases as discusse in the follwing sections.

1.1.2 Thermodynamics of Double Helix

The thermal stability of a generic DNA duplex depends on nitrogenous base sequence, quality of pairings and length. It can be evaluated as the change in free energy, ΔG, between the DNA duplex and the unbound strands. The variation of free energy is determined by a variation of both enthalpy, ΔH and entropyΔS. Assuming ΔH and ΔS as temperature (T) independent, the variation of free energy can be expressed as $\Delta G = \Delta H - T\Delta S$. The most

used model to evaluate such a contribution is the "Nearest Neighbour" (NN) model[8]. This model considers units composed by two consecutive base pairs in the duplex. Such a unit is called quartet and it naturally includes both base pairing and base stacking energies, that are hardly decoupled. There are 10 different possible quartets with corresponding specific thermodynamic parameters which are ΔG_{NN}, ΔH_{NN} and ΔS_{NN}, which are calculated in[8]. Thus the energy of a duplex is given by:

$$\Delta G(T) = \Delta G_{int}(T) + \sum \Delta G_{NN}(T) - 0.114(N/2)\ln[Na^+], \tag{1.1}$$

where ΔG_{int} is the energetic penalty due to duplex formation respectively, whereas the last term is the salt correction where N is the number of phosphates and $[Na^+]$ is the molar salt concentration. It's important to observe that the contribution due to penalty symmetry for self-complementary sequences is not considered here. Indeed, in[8] it has been introduced as an ad hoc fit parameter, whereas more recently in[9] this penalty contribution has been properly evaluated as entropic effect.
The thermal stability can be tested by increasing the temperature and consequently causing the break of the hydrogen bonds between base-pairs and leading to a complete denaturation of the double helix in single strands. Such a phenomena can be experimentally investigated by UV absorbance as thoroughly discussed in Chapter 3. The melting temperature T_m is defined as the temperature where half of DNA sequences are bound in double helix configuration and the remaining half are unbound, i.e. $\Delta G = 0$. Assuming that the transition is two-state, i.e. DNA sequences are either bound in double helix or unbound, the T_m at $c_{NaCl} = 1M$ is[9]:

$$T_m = \frac{\Delta H}{\Delta S + k_B \ln\left(\frac{[c]}{x}\right)}, \tag{1.2}$$

where [c] is the number concentration in mol/l, k_B is the Boltzmann constant, x=2 for self-complementary DNA sequences and x=4 for nonself-complementary DNA sequences.

1.2 DNA Liquid crystals

Solid, liquid or gas are one of the various states of matter, and they are all characterized by a specific positional and orientational order of the molecules they are composed by. For example in isotropic liquid phases, the particles have neither orientational nor positional order. To the contrary, solid phases can have both kind of orders over a long range. Liquid crystal is a state of matter in between these two classes, where particles show both positional and orientational order and they can move as in a fluid.

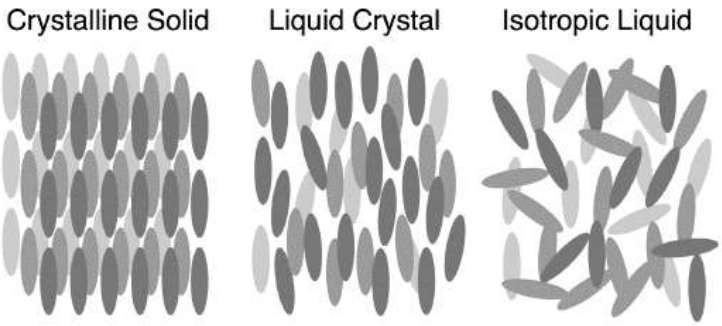

Figure 1.2: Three different states of matter. Figure adapted from[10].

1.2.1 Long nucleic acids

Experimental evidences that long DNA macromolecules can assembly in a LC phase have been observed since the discovery of the DNA helix itself, from the "photo 51", taken by Rosalind Franklin, whose diffraction pattern was indeed given by nematic liquid crystal state[11]. DNA LC were also found in living species such as bacteria and viruses, where the concentration of DNA can reach up to 800 mg/ml and support biological activity[12]. Among the several LC phases, in the following the two DNA LC mesophases involved in my thesis work are introduced.

Chiral Nematic LC phase

The nematic (N) phase is made of anisotropic molecules, e.g. rigid rods, that share an average alignment parallel to a unit vector $\boldsymbol{n}$. called director. States characterized by $\boldsymbol{n}$ and -$\boldsymbol{n}$ are equivalent. The order parameter that characterises the insurgence of nematic order is given by[13]:

$$S = \frac{1}{2}\langle 3\cos^2\theta - 1\rangle, \tag{1.3}$$

where θ is the angle between $\boldsymbol{n}$ and the molecular axis, and the statistical average is both spatial and temporal. When $S = 0$ there is no alignment ($\langle\cos^2\theta\rangle = 1/3$). To the contrary, S= 1 indicates perfect alignment($\langle\cos^2\theta\rangle = 1$). So, its variation from zero value indicates the insurgence of orientational order in a nematic phase and it quantifies the alignment.
Chiral molecules, like DNA, shows a cholesteric nematic (N^*) phase which is made of superimposed nematic layers twisted one over the other that give rise to a macroscopic helical superstructure, whose axis is always orthogonal to local director $\boldsymbol{n}$, as shown in Fig.1.3. The periodicity of such super-helix is $p/2$, where p is the cholesteric pitch whose size range from to 0.1μm to 10 μm. Fig.1.4 shows a polarized image of N^* LC of short DNA duplexes.

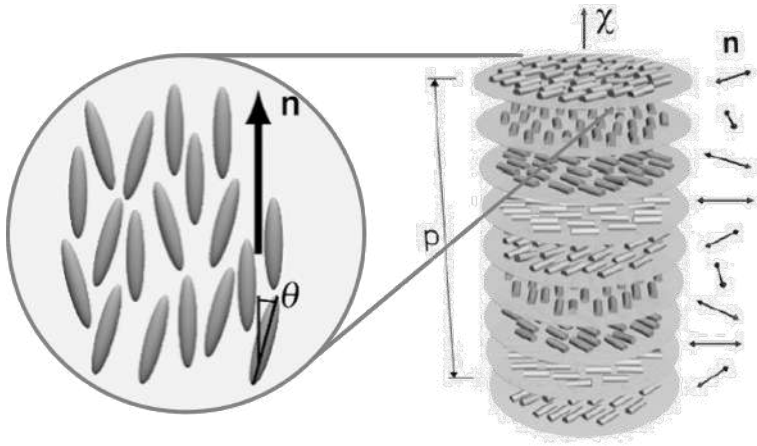

Figure 1.3: Panel left: representation of the nematic phase, with director $\boldsymbol{n}$ and the angle θ between $\boldsymbol{n}$ and optical axis of the molecule. Right panel: representation of the cholesteric phase given by superposition of nematic layers twisted one over the other around the super-helix axis χ. p is the cholesteric pitch of cholesteric phase.

Columnar LC phase

The columnar (COL) liquid crystal phase is made of stacked aggregates into long columns that form a 2D hexagonal lattice. Such columns can flow one over the other and they can bend giving rise to focal conics, the characteristic texture of columnar LC, as shown in Fig.1.4 and 1.5.

The transition of solution of long DNA molecules ($>$ 24 base pairs) in liquid crystal phases can be explained in the framework of Onsager's theory[14] as an entropic effect. For a solution of long hard rods at high enough concentration, the entropy is maximized by ordering the rods in a nematic phase where the positional entropy increase is higher than the decrease of orientational entropy. In other words, if the rods didn't get organized in nematic phase their movements would be more constrained (lower entropy). The Onsager prediction claims that the volume fraction at which the isotropic nematic transition occurs is $\phi_{IN} \approx 4D/L$, where D and L are the diameter and the length of the rod, respectively. Such theoretical prediction can be applied also to DNA duplexes. Although DNA duplexes are charged rods, they can be described as hard rods with an higher effective diameter that takes into account the repulsive interaction screened by ionic strength.

1.2.2 Short nucleic acids

Surprisingly, it has been experimentally observed that shorter DNA duplexes, from 20 to single base pairs long[15, 16, 17] can support LC phases. According

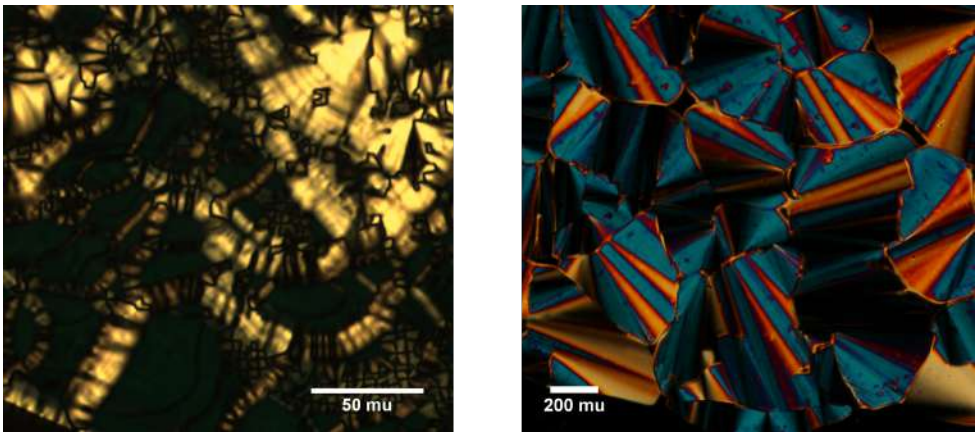

Figure 1.4: Polarized images of liquid crystal phases of short DNA duplexes, 12 nucleobases long. Left figure: cholesteric LC, green colour reveals cholesteric pitch p being ≈ 550nm. Right panel: columnar LC with typical focal conics.

to Onsager theory such a phenomena cannot occur for rods with $L/D < 4$. Actually the Onsager limit is still valid. Indeed, shorter blunt ended duplexes, in order to minimize the contact of paired nitrogenous bases with water molecules, pile one over the other thanks to stacking energies. So that longer and longer columnar aggregates are able to form until they reach the threshold length set by Onsager prediction that triggers the nematic transition. Further raise of the concentration leads to columnar phase, as shown in Fig.1.5.
This linear aggregation process can be easily suppressed by adding an unpaired base at both ends of a blunt-ended double helix. Differently, it can be enhanced by adding self complementary nucleobases at the terminals of a blunt-end duplex[18].
There are also other ways that favour the insurgence of a LC phase. For example polycations such as polylysine[19, 20], protamine and spermine (as it will be shown in Chapter6), when properly mixed with DNA duplexes in a wide range of ionic strength, they promote the insurgence of DNA LC. Such polycations introduce attractive inter-helical interactions that make possible LC phase even at lower DNA concentration.

1.2.3 Average length of DNA linear aggregates

Several theoretical models have been developed and Monte Carlo (MC) simulations performed to describe the average length growth of linear aggregates of short duplexes and corresponding onset of liquid crystal phase[21, 22].
In particular, De Michele et al.[22] performed MC simulation approximating short DNA duplexes with aspect ratio $X_0 = L/D$ (ranging from 1 to

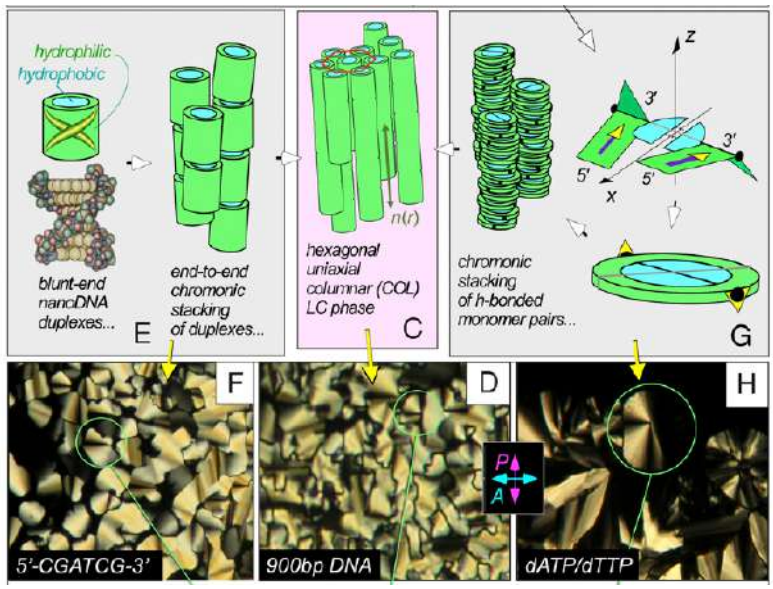

Figure 1.5: Self-assembly aggregation in COL liquid crystal phase of long DNA (900 base pairs long), short DNA duplexes (6 nucleobases long) and monomer pairs, all of them driven by base stacking along with base pairing. Figure adapted from[17].

3) as cylinders with attractive patches that mimic base-stacking interaction. By assuming the chain length distribution be exponential with the average chain length M, and minimizing the free energy suggested by Kuriabova et al. in[21] of such system, they found that M in the isotropic phase is:

$$M = \frac{1}{2}(1 + \sqrt{1 + 4e^{\sigma_b}\phi e^{k_I \phi \eta(\phi) + \beta \Delta H}}), \tag{1.4}$$

where ϕ is the volume fraction, $\beta = 1/k_B T$, k_I is a constant that depends on X_0, $\eta(\phi)$ is the Parson Lee factor[23], σ_b and ΔH are the entropic and the energetic contribution of the bonding, respectively. Fig.1.6 shows MC simulation (symbols) and theoretical predictions (dashed lines) of M dependence as a function of the volume fraction ϕ of short DNA duplexes with X_0=2 in both isotropic and nematic phase. In the isotropic phase M increases with ϕ and the lower the bonding energy ΔE_s the bigger the increment of M from the isotropic to nematic phase and the higher the ϕ at which isotropic-nematic transition occurs.

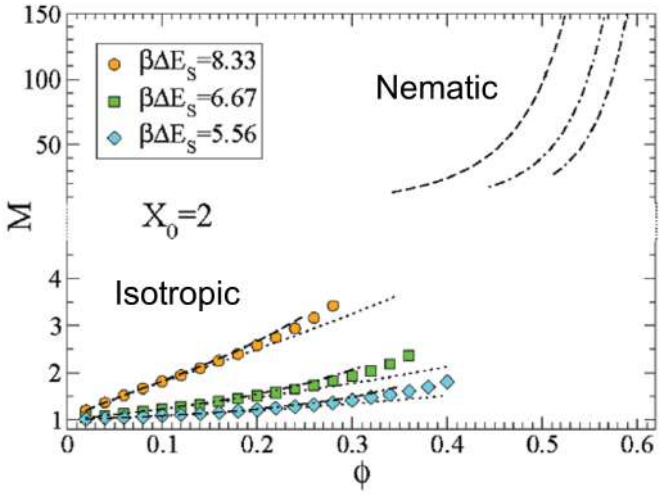

Figure 1.6: Aggregation number as a function of the volume fraction ϕ for cylinders with $X_0 = L/D = 2$ with attractive patches of different bonding energies as reported in the legend. Symbols: MC simulation results. Dashed lines: theoretical prediction in the isotropic phase (M<4) and nematic phase (M>20). Dotted lines: Onsager second viral approximation. Figure adapted from[22].

Bibliography

[1] Chris R Calladine and Horace Drew. *Understanding DNA: the molecule and how it works*. Academic press, 1997.

[2] Eric T Kool. Hydrogen bonding, base stacking, and steric effects in dna replication. *Annual review of biophysics and biomolecular structure*, 30 (1):1–22, 2001.

[3] Martin J Packer, Mark P Dauncey, and Christopher A Hunter. Sequence-dependent dna structure: dinucleotide conformational maps. *Journal of molecular biology*, 295(1):71–83, 2000.

[4] Jiří Šponer, Jerzy Leszczyński, and Pavel Hobza. Nature of nucleic acid-base stacking: nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. comparison of stacked and h-bonded base pairs. *The Journal of Physical Chemistry*, 100(13):5590–5596, 1996.

[5] Jiří Šponer, Petr Jurečka, Ivan Marchan, F Javier Luque, Modesto Orozco, and Pavel Hobza. Nature of base stacking: reference quantum-chemical stacking energies in ten unique b-dna base-pair steps. *Chemistry–A European Journal*, 12(10):2854–2865, 2006.

[6] Daniel Svozil, Pavel Hobza, and Jiri Sponer. Comparison of intrinsic stacking energies of ten unique dinucleotide steps in a-rna and b-dna duplexes. can we determine correct order of stability by quantum-chemical calculations? *The Journal of Physical Chemistry B*, 114(2):1191–1203, 2010.

[7] Richard Wing, Horace Drew, Tsunehiro Takano, Chris Broka, Shoji Tanaka, Keiichi Itakura, and Richard E Dickerson. Crystal structure analysis of a complete turn of b-dna. *Nature*, 287(5784):755–758, 1980.

[8] John SantaLucia Jr and Donald Hicks. The thermodynamics of dna structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440, 2004.

[9] Carlos A. Plata, Stefano Marni, Amos Maritan, Tommaso Bellini, and Samir Suweis. Statistical physics of dna hybridization. *Phys. Rev. E*, 103:042503, Apr 2021. doi: 10.1103/PhysRevE.103.042503. URL https://link.aps.org/doi/10.1103/PhysRevE.103.042503.

[10] Marco Todisco. *A Liquid Crystal World at the Origin of Life*. PhD Thesis, 2020.

[11] JE Lydon. The dna double helix—the untold story. *Liquid Crystals Today*, 12(2):1–9, 2003.

[12] Françoise Livolant. Ordered phases of dna in vivo and in vitro. *Physica A: Statistical Mechanics and its Applications*, 176(1):117–137, 1991.

[13] Pierre-Gilles De Gennes and Jacques Prost. *The physics of liquid crystals*. Number 83. Oxford university press, 1993.

[14] Lars Onsager. The effects of shape on the interaction of colloidal particles. *Annals of the New York Academy of Sciences*, 51(4):627–659, 1949.

[15] Michi Nakata, Giuliano Zanchetta, Brandon D Chapman, Christopher D Jones, Julie O Cross, Ronald Pindak, Tommaso Bellini, and Noel A Clark. End-to-end stacking and liquid crystal condensation of 6–to 20–base pair dna duplexes. *Science*, 318(5854):1276–1279, 2007.

[16] Tommaso P Fraccia, Gregory P Smith, Lucas Bethge, Giuliano Zanchetta, Giovanni Nava, Sven Klussmann, Noel A Clark, and Tommaso Bellini. Liquid crystal ordering and isotropic gelation in solutions of four-base-long dna oligomers. *ACS nano*, 10(9):8508–8516, 2016.

[17] Gregory P Smith, Tommaso P Fraccia, Marco Todisco, Giuliano Zanchetta, Chenhui Zhu, Emily Hayden, Tommaso Bellini, and Noel A Clark. Backbone-free duplex-stacked monomer nucleic acids exhibiting watson–crick selectivity. *Proceedings of the National Academy of Sciences*, 115(33):E7658–E7664, 2018.

[18] Simone Di Leo, Marco Todisco, Tommaso Bellini, and Tommaso P Fraccia. Phase separations, liquid crystal ordering and molecular partitioning in mixtures of peg and dna oligomers. *Liquid Crystals*, 45(13-15):2306–2318, 2018.

[19] J Pelta Jr, Dominique Durand, Jean Doucet, and Francoise Livolant. Dna mesophases induced by spermidine: structural properties and biological implications. *Biophysical journal*, 71(1):48–63, 1996.

[20] Tommaso P Fraccia and Tony Z Jia. Liquid crystal coacervates composed of short double-stranded dna and cationic peptides. *ACS nano*, 14(11): 15071–15082, 2020.

[21] Tatiana Kuriabova, MD Betterton, and Matthew A Glaser. Linear aggregation and liquid-crystalline order: comparison of monte carlo simulation and analytic theory. *Journal of Materials Chemistry*, 20(46): 10366–10383, 2010.

[22] Cristiano De Michele, Tommaso Bellini, and Francesco Sciortino. Self-assembly of bifunctional patchy particles with anisotropic shape into polymers chains: Theory, simulations, and experiments. *Macromolecules*, 45(2):1090–1106, 2012.

[23] JD Parsons. Nematic ordering in a system of rods. *Physical Review A*, 19(3):1225, 1979.

Chapter 2

Experimental Methods

2.1 Light Scattering

Light scattering is a powerful optical technique that gives the possibility to investigate noninvasively important properties of sample on the micron and sub-microns length scale. In particular, static light scattering experiments allows to measure static features as the mass and the size of growing clusters[1], whereas dynamic light scattering technique give access to dynamic features as diffusion coefficient and size of diffusing particles. The latter technique find application in various fields: in cosmetics, for example to measure the size of inorganic addictive in sunscreen[2]; in food, for example in limoncello, to detect the size of oil droplets[3] that give flavour to such a marvelous liquor; to detect indirectly the formation of structures as gel, by measuring the variation of a probe particle diffusion[4].
Both static and dynamic light scattering techniques are based on the ability of particles to diffuse the incoming light. Indeed, when an electromagnetic wave impinges on a particle, it emits a spherical electromagnetic wave, whose amplitude is proportional to the incident one. The inner structure of diffusing particles and their mutual interactions affect the diffused light.

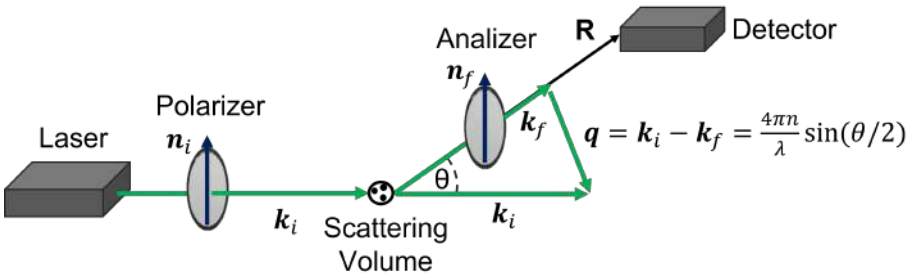

Figure 2.1: Schematic sketch of light scattering setup and wave vector q.

Such features can be discovered by looking and analysing the scattered light in a sort of a reverse process that is explained in the following.
Here only we are interested in Quasi-Elastic Light Scattering (QELS) processes, where multiple scattering events are negligible and the scattered filed E_s is proportional to the incoming field E_0. Fig.2.1 shows the sketch of a light scattering setup. Incident radiation of wavelength λ comes from a laser and its initial polarization, $\boldsymbol{n}_i$, can be set by a polarizer. The polarized light impinges on the sample, and it is scattered by particles in the scattering volume. It passes through an analyzer which select the finale polarization, $\boldsymbol{n}_f$, and it is collected by a detector at a distance $\boldsymbol{R}$. $\boldsymbol{k}_i$ and $\boldsymbol{k}_f$ are wavevectors of incoming and scattered beam, respectively. θ is the angle formed by the direction of $\boldsymbol{k}_i$ and $\boldsymbol{k}_f$. $\boldsymbol{q}$ is the transferred momentum defined as $\boldsymbol{q} = \boldsymbol{k}_i - \boldsymbol{k}_f$. Being the process quasi elastic, $|\boldsymbol{k}_i| \approx |\boldsymbol{k}_f| = \frac{2\pi n}{\lambda}$ (n is the refractive index of

sample) and consequently:

$$q = |\boldsymbol{q}| = 2k \sin \frac{\theta}{2} \tag{2.1}$$

2.1.1 Molecular approach to light scattering

As mentioned in the introduction, the particles of interest in this work are short DNA duplexes, few nanometers long. So, their dimension is small compared to the laser wavelength λ=532 nm and Rayleigh theory can be safely applied. DNA duplexes are anisotropic particles, so let's see firstly the field scattered by an anisotropic molecule characterized by a polarizability tensor $\boldsymbol{\alpha}$ and whose center of mass is described by $\boldsymbol{r}$(t). The incident field $\boldsymbol{E}_i$ is:

$$\boldsymbol{E}_i(\boldsymbol{R}, t) = \boldsymbol{n}_i E_0 e^{i(\boldsymbol{k}_i \cdot \boldsymbol{R} - \omega t)} \tag{2.2}$$

The dipole moment induced by a incident field $\boldsymbol{E}_i(\boldsymbol{R}$,t) is:

$$\boldsymbol{\mu}(t) = \boldsymbol{\alpha}(t) \cdot \boldsymbol{E}_i(\boldsymbol{R}, t). \tag{2.3}$$

The dipole moment varies with time and consequently emits electromagnetic radiation. The scattered field with final polarization $\boldsymbol{n}_f$, which is collected by the detector at a large distance $\boldsymbol{R}$, is:

$$\boldsymbol{E}_{if}(\boldsymbol{R}, t) = \frac{-k_f^2 E_0}{4\pi\epsilon R} e^{i(k_f R - \omega t)} e^{i\boldsymbol{q}\cdot\boldsymbol{r}(t)} \alpha_{if}(t), \tag{2.4}$$

where:

$$\alpha_{if}(t) = \boldsymbol{n}_f \cdot \boldsymbol{\alpha}(t) \cdot \boldsymbol{n}_i \tag{2.5}$$

is the component of the polarizability tensor along $\boldsymbol{n}_i$ and $\boldsymbol{n}_f$ directions. The time dependencies are due to the translation motion of molecule, $e^{i\boldsymbol{q}\cdot\boldsymbol{r}(t)}$, and to both vibration and rotation of the molecule, $\alpha_{if}(t)$.
If the electronic clouds of molecules are not perturbed by neighbouring molecules, the total scattered field will be simply proportional to the superposition of scattered fields given by eq2.4 from each molecule j irradiated in the scattering volume V:

$$E_{if}(\boldsymbol{q}, t) \propto \sum_{j \in V} \alpha_{if}^j(t) e^{i\boldsymbol{q}\cdot\boldsymbol{r}_j(t)}. \tag{2.6}$$

The experimental quantity that is accessible is the intensity $I_{if}(\boldsymbol{q}, t)$:

$$I_{if}(\boldsymbol{q}, t) = E_{if}^*(\boldsymbol{q}, t) E_{if}(\boldsymbol{q}, t) \propto \sum_{j,l \in V} \alpha_{if}^j(t) \alpha_{if}^l(t) e^{-i\boldsymbol{q}\cdot(\boldsymbol{r}_l(t) - \boldsymbol{r}_j(t)}. \tag{2.7}$$

The time average of eq.2.7 gives the average scattered intensity at the corresponding transferred momentum $\boldsymbol{q}$.

$$I_{if}(\boldsymbol{q}) = \langle E_{if}^*(\boldsymbol{q}, t) E_{if}(\boldsymbol{q}, t) \rangle \propto \langle \sum_{j,l \in V} \alpha_{if}^j(t) \alpha_{if}^l(t) e^{-i\boldsymbol{q}\cdot(\boldsymbol{r}_l(t) - \boldsymbol{r}_j(t)} \rangle. \tag{2.8}$$

2.1.2 Time scale of intensity fluctuations

Because of the motion of particles the scattered intensity fluctuates. Without entering in the details, in the following I show to get information on the diffusion of particles from such intensity fluctuations. Let's consider two particles 1 and 2 whose relative displacement at time t=0 is $\boldsymbol{r}_{12}(0) = \boldsymbol{r}_1(0) - \boldsymbol{r}_2(0)$. The corresponding intensity I(0) will be given by eq.2.7 where the scattered field is the superposition of the fields scattered by the two particles with a phase difference $\phi(0) = \boldsymbol{q} \cdot \boldsymbol{r_{12}}$. With time, such a phase difference will change because of the motion of the two particles. So, there is a critical time τ_c, such that for $t \leq \tau_c$ I(t) and I(0) slightly differs, i.e. the two intensities are correlated. Whereas for $t \geq \tau_c$ I(t) and I(0) are completely different (uncorrelated). This happens when:

$$\Delta\phi = \phi(\tau_c) - \phi_0 = \boldsymbol{q} \cdot (\boldsymbol{r_{12}}(\tau_c) - \boldsymbol{r_{12}}(0)) \approx 1. \tag{2.9}$$

The relative displacement $\Delta\boldsymbol{r}_{12}(\tau_c)$=$\boldsymbol{r_{12}}(\tau_c) - \boldsymbol{r_{12}}(0)$ will be of the order $|\Delta\boldsymbol{r}_{12}(\tau_c)| \approx \sqrt{D\tau_c}$, where D is te diffusion coefficient of particle, and consequently the time scale τ_c is:

$$\tau_c \approx \frac{1}{Dq^2}. \tag{2.10}$$

Therefore, the τ_c depends on the scattering angle θ, through eq.2.1, i.e. τ_c is faster for light scattered at high angles and slower for light scattered at low angles by the same particle. Such characteristic time can be measured in dynamic light scattering experiments from intensity (or field) correlation functions.

Diffusion coefficients and hydrodynamic radius

The translational diffusion coefficient D_T can be calculated by solving out the Langevin equation for a particle subjected to a random force and to a dissipative viscous force $\boldsymbol{F} = -f_t\boldsymbol{v}$, where v is the velocity of the particle and f_t is friction coefficient. For a sphere $f_t = 6\pi\eta R$, where η is the viscosity of solution, R is the radius of the particle. The solution of Langevin equation provides through the Einstein relation $D = k_BT/f$, the D_T as:

$$D_T = \frac{k_BT}{f_t} = \frac{k_BT}{6\pi\eta R}, \tag{2.11}$$

Einstein relation provides also the rotational diffusion coefficient, where the generic friction coefficient f is replaced by the rotational friction drag $f_{rot} = 8\pi\eta R^3$ providing:

$$D_{rot} = \frac{k_BT}{8\pi\eta R^3} \tag{2.12}$$

In real experiments, particles are surrounded by solvent molecules, so that the effective measured radius is bigger than the real one. Such radius is called hydrodynamic radius R_h, and whatever the shape of the particle, it can described as a sphere with equivalent volume V and radius R_h.

2.1.3 Correlation functions

The fluctuations of scattered fields can be evaluated by using the correlation function of scattered field normalized to its value at time t =0:

$$G_{1,if}(\tau) = \frac{\langle E^*_{if}(t)E_{if}(t+\tau)\rangle}{\langle E^*_{if}(0)E_{if}(0)\rangle}, \tag{2.13}$$

where $\langle E^*_{if}(0)E_{if}(0)\rangle$ is the average scattered intensity.

Dilute solution of anisotropic particles

Let's calculate, as an example, the field correlation function of a dilute solution of particle with anisotropic polarizability $\boldsymbol{\alpha}$, whose scattered fields are given by eq.2.4. Because the solution is diluted, only self correlation must be considered, and consequently according to eq.2.13 with t=0, the field correlation function is:

$$G_{1,if}(\tau) = \sum_{j=1}^{N} \langle \alpha^j_{if}(0)\alpha^j_{if}(\tau)e^{i\boldsymbol{q}\cdot(\boldsymbol{r}_j(\tau)-\boldsymbol{r}_j(0)}\rangle, \tag{2.14}$$

where N is the number of scatterers in the scattering volume.
If translation and rotation are statistical independent, eq.2.14 can be further simplified in[5]:

$$G_{1,if}(\tau) = \langle N\rangle\langle \alpha_{if}(0)\alpha_{if}(\tau)\rangle e^{-Dq^2\tau} \tag{2.15}$$

Siegert Relation

Experimentally, it is not possible to measure the scattered and consequently calculate the field correlation function. Instead it is possible to measure the scattered intensity and calculate the intensity field correlation function defined as:

$$G_{2,if}(\tau) = \frac{\langle I_{if}(t)I_{if}(t+\tau)\rangle}{\langle I\rangle^2}, \tag{2.16}$$

where the denominator normalize the intenisty field correlation for $\langle I\rangle^2$, which corresponds to completely uncorrelated intensity signals, such that $G_2(\tau =$

$\infty) = 1$. Field and intensity field correlation functions are related by the Siegert relation:

$$G_{2,if}(\tau) = 1 + |G_{1,if}(\tau)|^2, \tag{2.17}$$

For the simple case of a dilute solution of isotropic noninteracting spherical particles, it can be shown[5] that $G_{2,if}(\tau) = 1 + e^{-2D_T q^2 \tau}$. From the decay time, $\tau_c = 1/(2D_T q^2)$, of the correlation function $G_2(\tau)$, it is possible to calculate the diffusion coefficient as $D_T = 1/(2q^2\tau_c)$ and consequently the hydrodynamic radius from eq.2.11. For this reason DLS is a really useful technique, because it allows to measure in a noninvasive way particle of microns and submicron size, that are not measurable and observable through video microscopy.

2.1.4 Polidispersity characterized by stretching exponent

Samples analysed in DLS experiments could reveal a polidisperse solution, due to various causes such as deterioration and impurities of the sample, or as in our specific case due to aggregation process that lead to polidisperse size clusters. The latter case is physical meaningful. A variety of sizes leads to a widening of characteristic times τ_c that can be described by[6]:

$$|G_1(\tau)| = e^{-\left(\frac{\tau}{\tau_c}\right)^{\alpha}}. \tag{2.18}$$

Let's define $\langle\tau_c\rangle$ the integral of the stretching exponential on the positive semi-axis[7]:

$$\langle\tau_c\rangle = \int_0^{\infty} dt e^{-\left(\frac{\tau}{\tau_c}\right)^{\alpha}} = \frac{\tau_c}{\alpha}\Gamma\left(\frac{1}{\alpha}\right), \tag{2.19}$$

where Γ is the gamma function.

2.1.5 Experimental Static Light Scattering Setup

The experimental apparatus, sketched in Fig.2.2, that allows to perform static depolarized light scattering (SDLS) measurements is a home-made setup, and it is assembled with the following components: a Nd:YAG double CV laser with emitting light at $\lambda = 532$ nm and power of the beam P = 300 mW; two polarizers located before and after the scattered light with horizontal and vertical polarization, respectively; a thermostat cylindrical chamber, whose temperature is regulated by a PID controller within 0.1 °C precision; a smaller cylinder containing the capillary with DNA sample (half of such cylinder is filled with filtered water to optimize sample thermalization); detection optics is composed by optical lens that make the scattered beam converge into large-core optical fibers with a inner core of 50 μm that guarantee a high spatial average (more than 100 speckles collected); fibers are plugged into the

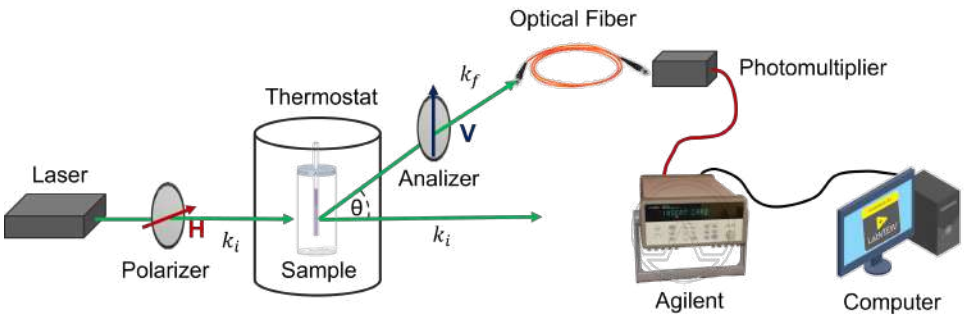

Figure 2.2: Experimental setup for static light scattering experiments.

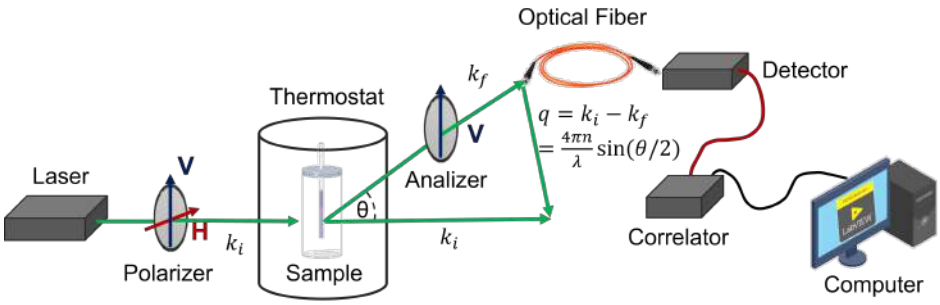

Figure 2.3: Experimental setup for dynamic light scattering experiments.

photomoltiplier tubes Hamamatsu 6780-20 (for silica photodiodes scattered signals are too low to be detected); Agilent A34970 data acquisition unit that converts the current signal into voltage signal; computer that allows through Labview software to control the temperature and record the voltage signal of scattered light. The scattered intensities are collected simultaneously at the angles $\theta = 90°$ and $135°$ through two different photomultipliers.

2.1.6 Experimental Dynamic Light Scattering Setup

The experimental apparatus, sketched in Fig.2.3, that allows to perform DLS measurements is a home-made setup, and it differs by SLS setup only in polarizer, which can be set also in vertical direction, and in the detection. In particular, the optical fibers are plugged into a photomultiplier tube, then an operational amplifier amplifies the signa into a digital one. Correlation functions are calculated by the four channel digital correlator $Flex03LQ-1$. The shortest τ that such correlator can calculate is 1.6×10^{-7}s; $Flex03LQ-1$ is plugged into computer that allows through Labview software to control the temperature and show the intensity field correlation functions. The scattered light is collected at the angle $\theta = 90°$.

2.2 FRAP

In this section, I briefly introduce diffusion phenomenon and its relation to Fluorescence Recovery After Photobleaching (FRAP) technique, that allows to measure the diffusion coefficient of a fluorescent molecule freely diffusing in solution. It's possible to bind a fluorescent marker to a target molecule. For example, fluorescein can be linked to a nucleobases of the DNA duplex (typically the thymine). In particular, in FRAP experiments a portion of the sample is overexposed to light, causing the photobleaching of fluorescent markers. After turning off the excitation light, a black spot appears. Over time, black spot turn gradually fluorescent because the diffusion of both fluorescent and photobleached markers, as qualitatively shown in Fig.2.5. So, by looking at recovery of fluorescence signal, the diffusion coefficient of fluorescent marker can be recovered.
The diffusion equation of particles moving of Brownian motion is[8]:

$$\frac{\partial \rho}{\partial t} = D\nabla^2 \rho, \tag{2.20}$$

where D is the diffusion coefficient, $\rho(\boldsymbol{r}, t)$ is the spatial distribution probability of particles located in $\boldsymbol{r}$ at time t. The solution in 2d is a Gaussian for $\rho(\boldsymbol{r}, t)$ that widens in time as $\sigma^2 = 2Dt$:

$$\rho(r, t) = \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \tag{2.21}$$

where $r = \sqrt{x^2 + y^2}$ is the modulus of position of the particle in polar coordinates. Therefore the amplitude of Gaussian distribution decreases with time t and the width broadens more and more as nicely shown in Fig.2.4.

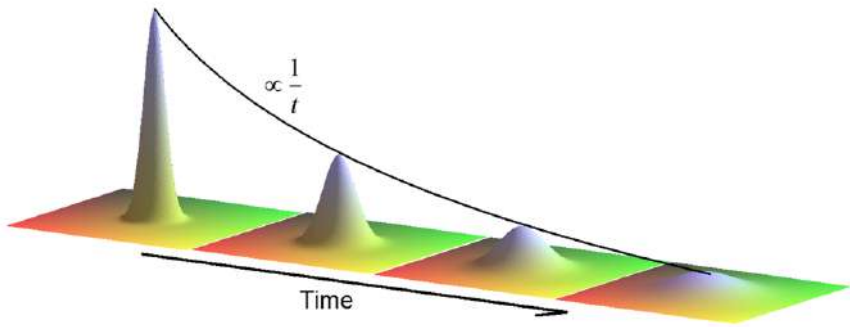

Figure 2.4: Evolution of the Gaussian distribution probability of diffusing particle by Brownian motion. Amplitude and and width in time are described by eq.2.21.

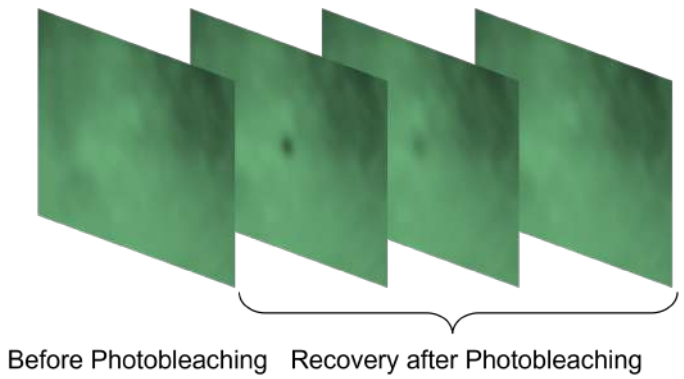

Figure 2.5: Qualitative representation of images sequence collective in FRAP experiment. Initial fluorescent sample (first image) is successively photobleached (balck spot in second image). Last two images show the recovery of fluorescence.

The calculation of diffusion coefficient has been performed as follows: firstly, image series were processed by subtracting a pre-bleached image; secondly the bleached spot center is localized in the first frame by fitting a 2D Gaussian function; thirdly, from such spot centre a radial average is performed on each image, so that it's possible to get the average signal, $F(r)$, as a function of distance r. Finally, $F(r)$was fitted by:

$$y_F = q + Ae^{-\frac{r^2}{2\sigma^2}}, \tag{2.22}$$

where q, A and σ are free parameters and represent plateau value, amplitude and width of $F(r)$. From the fits of each image at corresponding time t, it's therefore possibile to get the width σ^2 as a function of time and consequently the diffusion coefficient D from the slope of the linear fit of σ^2 as function of time t.

2.2.1 Experimental Setup

FRAP experiments were performed by means of the microscope Nikon TE300. Light source is the pe-4000 CoolLED lamp and λ=470nm was selected to excite fluorescein. Fluorescein bleaching were performed by focusing the light on a sample area with a beam waist ranging from $\approx 100\mu m$ to $\approx 800\mu m$, through a high magnification objective (20x or 50x) with standard Fluorescein filter cube from Nikon. Images are collected by a Nikon DS-Fi3 camera with a magnification 10x. Camera is synchronized with coolLED lamp to take pictures only when sample is illuminated by the lamp to reduce further photobleaching in the running experiment.
Sample are prepared in a 20 μm thickness glass microscope cells, which is located on a thermalized chamber, equipped with a peltier system.

Bibliography

[1] Marina Carpineti, F Ferri, Marzio Giglio, E Paganini, and U Perini. Salt-induced fast aggregation of polystyrene latex. *Physical Review A*, 42(12): 7347, 1990.

[2] Rosalynn Quiñones, Sara Moreno, Deben Shoup, Mieke Klein, Tamara D Westfall, and Aakriti Damai. Examining particle size of inorganic active ingredients within sunscreens using dynamic light scattering. *Journal of Chemical Education*, 98(4):1371–1380, 2021.

[3] Leonardo Chiappisi and Isabelle Grillo. Looking into limoncello: the structure of the italian liquor revealed by small-angle neutron scattering. *ACS omega*, 3(11):15407–15415, 2018.

[4] F Madonia, PL San Biagio, MU Palma, G Schiliro, S Musumeci, and G Russo. Photon scattering as a probe of microviscosity and channel size in gels such as sickle haemoglobin. *Nature*, 302(5907):412–415, 1983.

[5] Bruce J Berne and Robert Pecora. *Dynamic light scattering: with applications to chemistry, biology, and physics*. Courier Corporation, 2000.

[6] Tommaso Bellini. *Statica e dinamica della birifrangenza elettrica di liquidi complessi*. PhD Thesis, 1990.

[7] CP Lindsey and GD Patterson. Detailed comparison of the williams–watts and cole–davidson functions. *The Journal of chemical physics*, 73 (7):3348–3357, 1980.

[8] Albert Einstein. *Investigations on the Theory of the Brownian Movement*. Courier Corporation, 1956.

Chapter 3

Melting and Pairing Statistics of Random DNA Oligomers

3.1 Introduction

Several biological processes, such as miRNA regulation and PCR, are allowed by the selectivity and specificity of mutual interactions of the involved biomolecules. In these processes such interactions occur after countless collisions between partner molecules in a "superdiverse" environment, until the most energetic and stable pairing takes place [1]. Because of variety of possible interaction and the crowded environment in which they occur, the success of the best pairing possible depends also on biomolecules concentration, the entire network of pair interactions and their degeneracy and cooperativity. Although the intricacy of mutual interactions in crowded environment has been studied[2, 3], the statistical physics and thermodynamics and corresponding theoretical models have not been reported yet.
One of the biological systems that naturally allow to investigate such a selectivity in superdiverse environment are random sequence DNA (rsDNA) oligomers mixtures. rsDNA oligomers are DNA oligonucleotides where each position along its sequence is occupied by one out of the four canonical bases (Adenine, A, Tymine T, Cytosine C, Guanoisne G) with equal probability, as shown in Fig. 3.1(a). In this thesis I worked with rsDNA with length L = 8, 12 and 20. Therefore, in a rsDNA mixture of a specific length there are 4^L distinct sequences that nicely mimic the superdiverse context in which paring errors distribution, interaction selectivity and equilibration features can be investigated as a function of temperature.
DNA and RNA hybridization has been deeply investigated leading to several models to calculate and predict melting temperature and free energy of duplexes from the the single acid nucleic strands involved in the pairing [4, 5, 6]. The most used model to calculate free energy of duplexes is the Nearest Neighbour (NN) model. The calculation consists on the summation of elemental contributions, quartets. Such contributions are calculated from melting temperatures of DNA sequences and naturally consider both Watson and Crick (WC) and non-canonical pairings [4]. Such database are unfortunately not capable of describing rsDNA mixtures. Indeed, rsDNA oligomer can interact at least in $4^L \times 4^L$ different ways, as represented in Fig. 3.1(b). The hybridization state with highest probability have few bases paired consecutively (3 for L=12) as shown in the last example of Fig. 3.1(c). Other examples of pairings, that occur with lower probability, are represented in Fig. 3.1(c), ranging from perfect duplexes to duplexes with shift, external and internal mismatches. Each of these duplexes has a specific binding free energy that can be calculated with NN model, whereas the extreme case of two oligomers not complementary at all, as one made of only As and the other made of only Gs, has no binding free energy. For a rsDNA mixture of length L is not trivial to predict the statistics of the different hybridization states. Indeed, the heterogeneity of pairings increases with the number of permitted

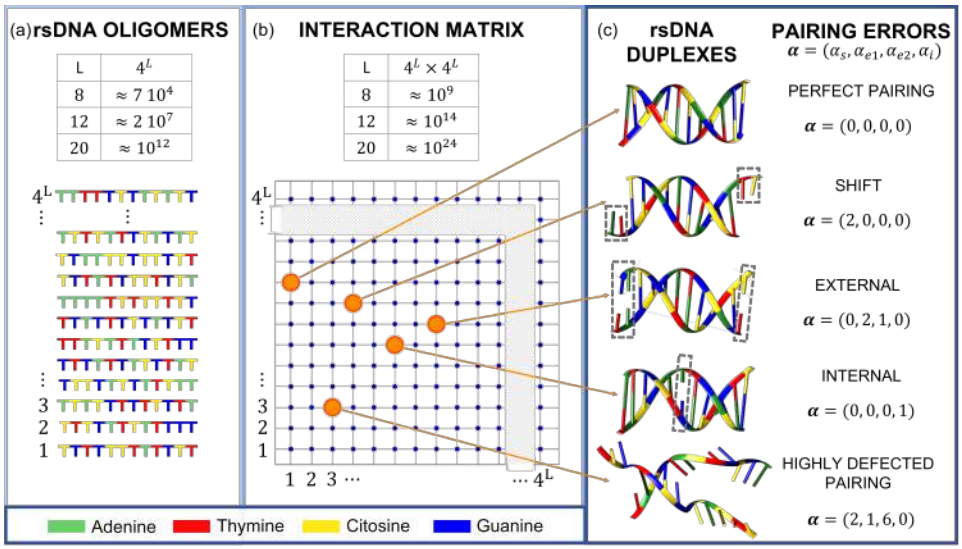

Figure 3.1: (a) Solutions of random-sequence DNA (rsDNA) oligomers of length L are mixtures made of 4^L distinct molecules, obtained by all the combinations of the four nucleobases, which are present at any position in the sequence with equal probability. (b) Each rsDNA oligomer can interact with 4^L different rsDNA oligomers, leading to a $4^L \times 4^L$ interaction matrix. Each dot in the matrix represents the most energetically favorable pairing between the two selected rsDNA oligomers, among all the possible mutual shifts. (c) Each position in the interaction matrix corresponds to a specific duplex motif, characterized by pairing errors which are here described by the parameters in $\boldsymbol{\alpha}$. The most probable duplex in the matrix is highly defected, as the last example in the panel.

pairing errors, whereas the corresponding binding energy drops. The competition between these two ingredients, degeneracy and binding free energy, are expected to set the rules for pairing statistics. In this chapter, I show the comprehension of selectivity and paring statistics in rsDNA mixtures by a combination of experimental technique and the development of a suitable theoretical framework.

3.2 Ensemble melting of rsDNA

The variation of DNA hyperchromicity, due to the melting of the DNA duplexes and detectable by UV light, allows to have an insight on the overall extent of hybridization in rsDNA mixtures. Fig. 3.2 shows the experimental melting curve $\theta_e(T)$ for 12N at $c_{rsDNA} = 0.04g/l$, $c_{NaCl} = 1M$ (green dots), compared with two reference $\theta_e(T)$ curves both at $c_{NaCl} = 1M$:blue and red dashed line corresponds to the melting curves, computed with standard tools,

of a solution composed by two complementary DNA 12mers with average ΔG (calculated according to Eq. 3.29, with $f_{CG} = 0.5$) at $c_{DNA} = 0.02\ g/l$ and $c_{DNA} = 0.04/4^{12}\ g/l$, respectively. Exactly, red dashed line is equivalent to the melting curve of a random DNA solution where only perfect duplexes are formed. It's evident the intermediate behaviour of θ_e of 12N, whose melting occurs in between the two dashed lines. In particular, rsDNA duplexes are both $\approx 30°C$ less stable than DNA duplexes formed by complementary strands at the same total concentration, and they are $\approx 5°C$ more stable than the 12mers complementary duplexes at the same concentration of rsDNA replicas. Moreover, the slope of $\theta_e(T)$ for 12N is clearly slighter than the other two slopes. Such observations reveal the formation of a variety of defected duplexes.
The same behaviour described for 12N occurs also for rsDNA with both higher and lower L. Fig.3.3 shows experimental $\theta_e(T)$ for 20N at $c_{rsDNA} = 0.04g/l$, $c_{NaCl} = 450mM$ compared with binary mixtures of 20mers at $c_{DNA} = 0.02\ g/l$ (blue dashed line) and $c_{DNA} = 0.04/4^{12}\ g/l$ (red dashed line), both at $c_{NaCl} = 450mM$. For 20N is reported only the first heating ramp as discussed in section 3.4.1. Fig.3.4 displays experimental $\theta_e(T)$ for 8N at $c_{rsDNA} = 25g/l$, $c_{NaCl} = 1M$ compared with binary mixtures of 8mers at $c_{DNA} = 12.5\ g/l$ (blue dashed line) and $c_{DNA} = 25/4^{12}\ g/l$ (red dashed line), both at $c_{NaCl} = 1M$.
These results gives us an initial important understanding on the selectivity of rsDNA pairings and suggest that such rsDNA behaviour could be independent on L, ionic strength and c_{rsDNA}. Fig. 3.5 clearly confirms such a hypothesis. In particular, Fig. 3.5(a) shows $\theta_e(T)$ at low concentration $c_{rsDNA} = 0.04$g/l for both 12N and 20N for three different salt concentration c_{NaCl}= [.15, .45, 1]M. Since the shorter the length L, the lower the T_m, to observe properly the melting for 8N it was necessary to increase its c_{rsDNA}. Such a achievement was not trivial at all and it is thoroughly described in section 3.5.3. Fig. 3.5(b) displays $\theta_e(T)$ at high concentration $c_{rsDNA} = 25$g/l for 8N for three different salt concentration c_{NaCl}= [.15, .45, 1]M. Fig 3.6 nicely sum up all these experiments, showing how T_m increases with L, c_{rsDNA} and logarithmically with c_{NaCl}, in agreement with literature [4, 7].

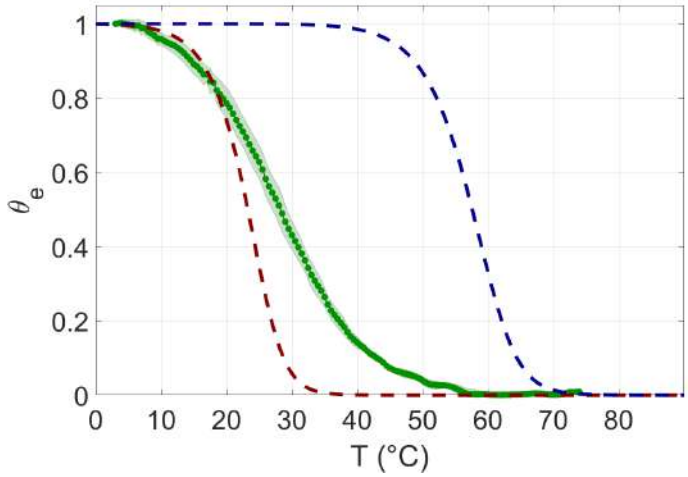

Figure 3.2: Ensemble melting curve as a function of temperature. Green dots: measured ensemble melting of 12N at c_{rsDNA}= 0.04 g/l. Shading marks experimental uncertainty resulting from the average over 8 experimental replicas. Dashed lines: theoretical melting predicted for equimolar solutions of two complementary 12mers at c_{DNA} = 0.02 g/l each (dashed blue line) and $c_{DNA} = 0.04/4^{12} \approx 2.4\ 10^{-9}$ g/l each (dashed red line). 12N $\theta_e(T)$ exhibits a behaviour intermediate between the two. c_{NaCl} = 1M in all curves.

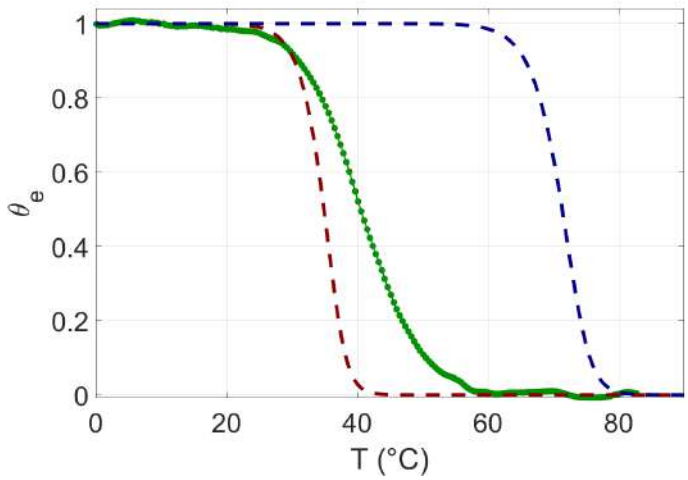

Figure 3.3: Ensemble melting curve as a function of temperature. Green dots: measured ensemble melting of 20N at c_{rsDNA}= 0.04 g/l. Dashed lines: theoretical melting predicted for equimolar solutions of two complementary 20mers at c_{DNA} = 0.02 g/l each (dashed blue line) and $c_{DNA} = 0.04/4^{20} \approx 3.6\ 10^{-14}$ g/l each (dashed red line). 20N $\theta_e(T)$ exhibits a behaviour intermediate between the two. c_{NaCl} = 1M in all curves.

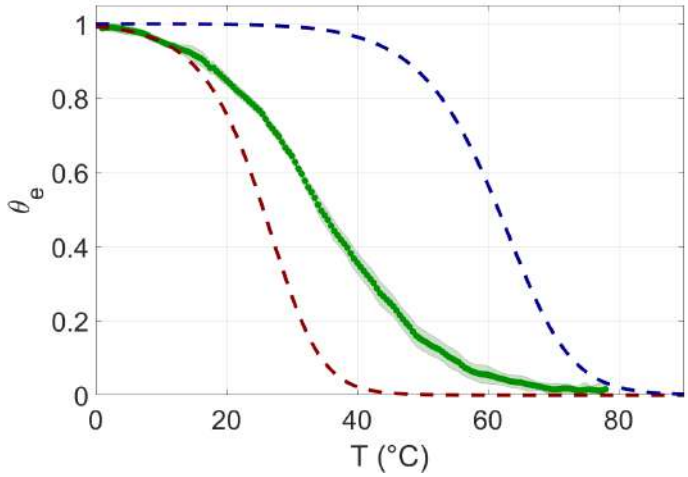

Figure 3.4: Ensemble melting curve as a function of temperature. Green dots: measured ensemble melting of 8N at c_{rsDNA}= 25 g/l. Shading marks experimental uncertainty resulting from the average over 8 experimental replicas. Dashed lines: theoretical melting predicted for equimolar solutions of two complementary 8mers at c_{DNA} =12.5 g/l each (dashed blue line) and $c_{DNA} = 25/4^8 \approx 3.8\ 10^{-4}$ g/l each (dashed red line). 8N $\theta_e(T)$ exhibits a behavior intermediate between the two. c_{NaCl} = 450mM in all curves.

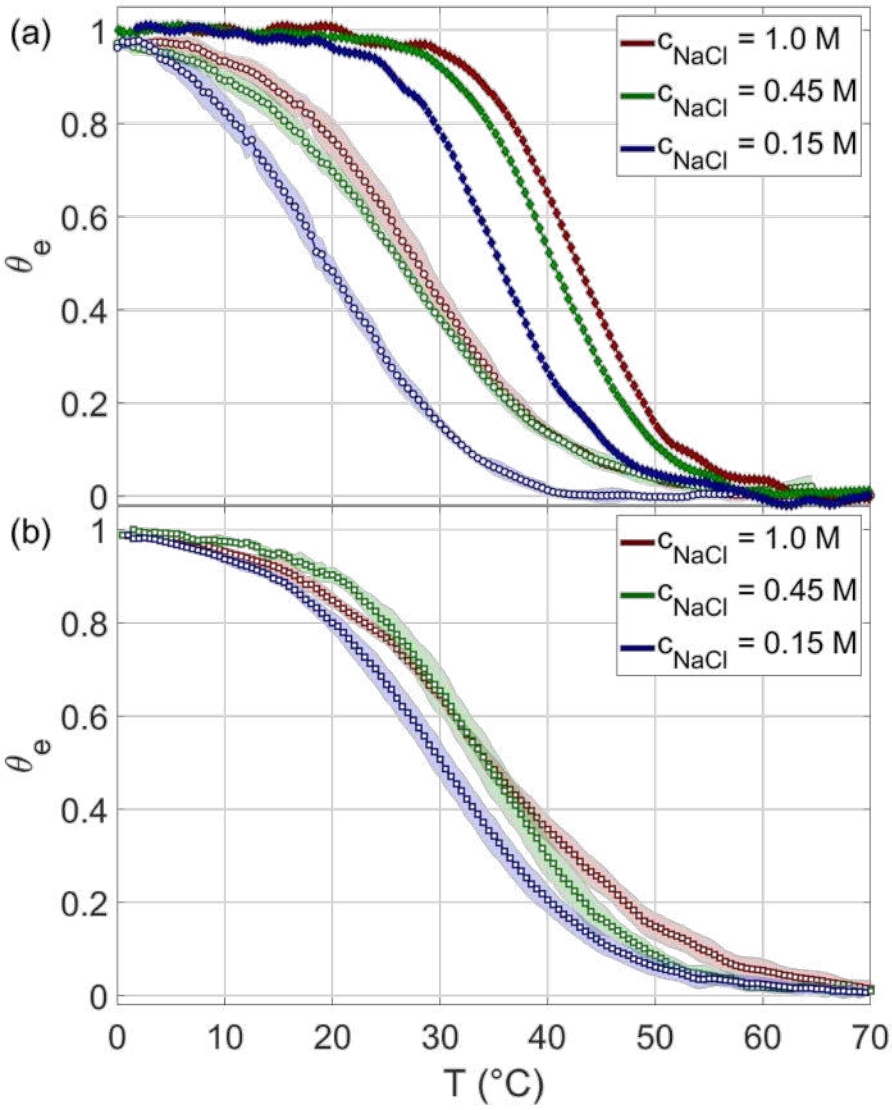

Figure 3.5: Experimental ensemble melting curves of rsDNA at three salt concentrations: c_{NaCl} = 1M (red) c_{NaCl} = 0.45M (green) and c_{NaCl} = 0.15M (blue). (a)Low concentration c_{rsDNA} = 0.04, for both 12N (open circles) and 20N (full diamonds). (b) High concentration c_{rsDNA} = 25 g/L for 8N (open squares). Shading marks experimental uncertainty resulting from the average over 6-8 experimental replicas for 8N and 12N, whereas, for 20N, just one experiment is shown as described in section3.4.1.

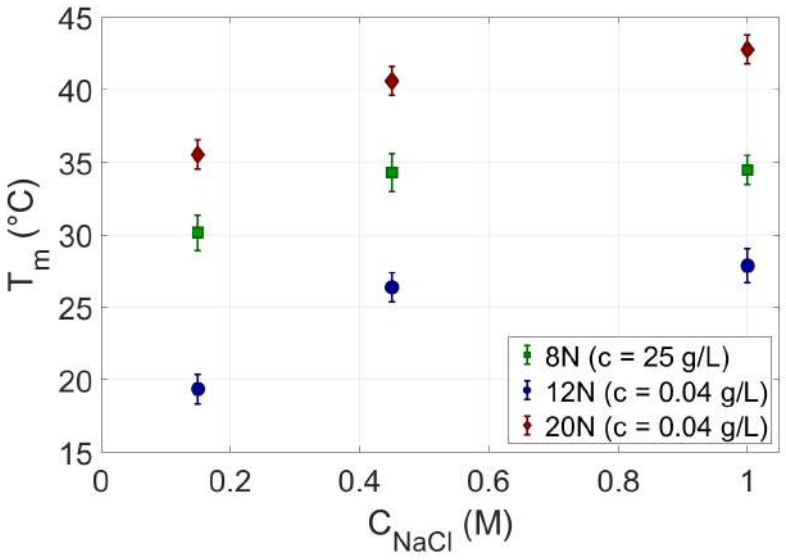

Figure 3.6: Experimental T_m for rsDNA solutions for 8N (green squares), 12N (blue circles) and 20N (red diamonds) obtained from corresponding melting curves (Fig. 3.5), as function of salt, c_{NaCl}.

3.3 Theoretical Framework

The development of the theoretical model that allows to predict rsDNA behaviour in melting experiments is here explained. Before describing in detail rsDNA mixture, in the first subsection the theoretical framework is introduced for a simpler case, where rsDNA oligomers are assumed to pair each other only with one specific binding energy[8].

3.3.1 Partition function for ideal rsDNA solution

For oligomers of equal length L, a rsDNA solution is composed by 4^L different equimolar sequences. Let's label these sequences from 1 to S and define N_i as the total number of the specific i-th sequence. So, the total number of sequences in the solution is $N = \sum_{i=1}^{S} N_i$ and the total concentration is $c = N/V$. The concentration of the i-th oligomer is $c_i = N_i/V = \chi_i c$, where χ_i is the fraction of sequences i-th, $\chi_i = N_i/N$, satisfying the normalization condition $\sum_{i=1}^{S} \chi_i = 1$. A generic i-th oligomer can pair with an other j-th, leading to a change of free energy $\Delta\tilde{G}_{ij} = G_{p,ij} - G_i - G_j$, where $G_{p,ij}$, G_i, G_j are the free energy of the duplex, of the ith and j-th oligomer, respectively. Let's assume that the oligomers can pair only with the most stable and energetic $G_{p,ij}$. When the ith and j-th oligomers pair each other, the volume that each oligomer can explore is reduced to a volume $V_{0,ij}$.

The goal is to get the partition function of our system in order to be able to calculate analytical expression of quantities such as the average number of duplexes formed by oligomers i-j-th, $\langle n_{ij} \rangle$.
The first step consists on evaluating the probability $p(\boldsymbol{n})$ of having a specific configuration of duplexes described by the vector $\boldsymbol{n} = (n_{11}, n_{12}, n_{13}, ..., n_{ij}, ..., n_{SS})$. Since the system is contact with a thermal bath at temperature T, the $p(\boldsymbol{n})$ is provided by canonical distribution:

$$p(\boldsymbol{n}) \propto f(\boldsymbol{n}) exp\Big\{-\beta\Big[\sum_i (N_i - 2n_{ii} - \sum_{i \neq j} n_{ij}) G_i + \sum_i \sum_{j \geq i} n_{ij} G_{p,ij}\Big]\Big\} \int d\boldsymbol{r_1} ... d\boldsymbol{r_N} d\boldsymbol{p_1} ... d\boldsymbol{p_N} exp\Big\{-\beta \sum_{i=1}^{N} \frac{\boldsymbol{p_i} \cdot \boldsymbol{p_i}}{2m}\Big\}, \tag{3.1}$$

where $f(\boldsymbol{n})$ counts the possible different configurations with exactly $\boldsymbol{n}$ paired duplexes (with identical total number of oligomers N) and $\beta = (k_B T)^{-1}$. The first exponential describes the inner degree of freedom of free and paired oligomers, whereas the integral over the $6N$-dimensional domain, describes the phase space explored by oligomers. To understand the contribution of free oligomers in first exponential term, let's consider the oligomer of species 1. The number of free oligomers of species 1 is the total number of oligomers of species 1, N_1, minus two times the number of duplexes formed by two 1-th oligomers, n_{11}, minus the number of duplexes $n_1 j$ ($j \neq 1$) in which only one 1-th oligomer is involved in the pairing.
Let's calculate now the integral and focus on the volume that oligomers can explore. Free oligomers and oligomers i paired with oligomers j in a duplex can explore all the volume V, so their spatial integral is $V^{N - n_{ij}}$. Whereas the j oligomer (paired to i oligomer) can explore only the volume $V_{0,ij}$ around the i oligomer, so its spatial integral is $V_{0,ij}^{n_{ij}}$. Therefore the integral result is:

$$\int d\boldsymbol{r_1} ... d\boldsymbol{r_N} d\boldsymbol{p_1} ... d\boldsymbol{p_N} exp\Big\{-\beta \sum_{i=1}^{N} \frac{\boldsymbol{p_i} \cdot \boldsymbol{p_i}}{2m}\Big\} = \Big(\frac{2\pi m}{\beta}\Big)^{\frac{3N}{2}} V^N \prod_i \prod_{j \geq i} \Big(\frac{V_{0,ij}}{V}\Big)^{n_{ij}}. \tag{3.2}$$

So, Eq.3.1 can written as:

$$p(\boldsymbol{n}) = \frac{1}{Z} f(\boldsymbol{n}) \prod_i \prod_{j \geq i} \Big[\frac{V_{0,ij}}{V} exp(-\beta \Delta \tilde{G}_{ij})\Big]^{n_{ij}}, \tag{3.3}$$

where in the partition function Z the constant factors are absorbed in. It's possible and convenient to write Eq.3.3 in a more compact and elegant way. Firstly, the volume $V_{0,ij}$ can be described as an entropic contribution $k_B \ln(N_A[V_{0,ij}])$, where the brackets denotes that unit of measurement for volume is litre. So $\Delta \tilde{G}_{ij} = \Delta \tilde{H}_{ij} - T \Delta \tilde{S}_{ij}$ turns to be $\Delta G_{ij} = \Delta \tilde{H}_{ij} - T(\Delta \tilde{S}_{ij} + k_B \ln(N_A[V_{0,ij}]))$.

Secondly, we can introduce the total concentration of oligomers, $c = N/V$, multiplying and dividing by N the prefactor of the exponential. So, Eq.3.3 can be written as:

$$p(\boldsymbol{n}) = \frac{1}{Z} f(\boldsymbol{n}) \prod_i \prod_{j \geq i} \left(\frac{\zeta_{ij}}{N} \right)^{n_{ij}}, \tag{3.4}$$

with the Boltzmann factor ζ_{ij} given by:

$$\zeta_{ij} = [c] exp(-\beta \Delta G_{ij}), \tag{3.5}$$

where brackets denote that concentration c is measured in mol/L. If ΔG is expressed in *cal*, the k_B must be substituted for the gas constant R. So, the partition function is:

$$Z = \sum_{\boldsymbol{n}} f(\boldsymbol{n}) \prod_i \prod_{j \geq i} \left(\frac{\zeta_{ij}}{N} \right)^{n_{ij}}, \tag{3.6}$$

where the sum is over all possible values n_{ij} satisfies the condition $0 \leq 2n_{ii} + \sum_{j \neq i} n_{ij} \leq N_i, \forall i$. The last term to calculate is $f(\boldsymbol{n})$. It can be calculated through combinatorics calculation as:

$$f(\boldsymbol{n}) = \prod_{i=1}^{S} \frac{N_i!}{2^{n_{ii}} (N_i - 2n_{ii} - \sum_{j \neq i} n_{ij})!} \prod_{j=i} \frac{1}{n_{ij}!}. \tag{3.7}$$

Let's consdier the i-th species. $N_i!$ counts all possible ways to order sequences of species i-th. Such term must be divided by : 1) factorial of the number of free oligomers of species i-th, $(N_i - 2n_{ii} - \sum_{j \neq i} n_{ij})!$ because they are indistinguishable each other; 2) $2^{n_{ii}}$ that takes into account that two identical oligomers are indistinguishable when paired, as well; 3) $\prod_{j=i} n_{ij}!$, i.e. all the configurations $\boldsymbol{n}$ are equivalent upon exchanging the duplexes n_{ij}, whatever j-th species.
Finally, it's possibile to evaluate the average number of duplexes n_{ij} form the partition function Z:

$$\langle n_{ij} \rangle = \zeta_{ij} \frac{\partial \ln Z}{\partial \zeta_{ij}}. \tag{3.8}$$

3.3.2 Partition function for real rsDNA mixture

Let's now consider properly rsDNA mixture. Differently from the previous case, the binding between two oligomers i-th and j-th can occur according to several modes $k = 1,..,K_{ij}$ (as sketched in Fig.3.1(c)), corresponding to global free energies change $\Delta G_{ij}^{(k)}$. Consequently the dimension of vector $\boldsymbol{n}$ is highly increased and corresponds to $\sum_{i=1}^{S} \sum_{j=1}^{S} K_{ij}$. The components of $\boldsymbol{n}$ are now $n_{ij}^{(k)}$ which count the duplexes formed by i-th and j-th according to

k-th mode. Following the same reasoning of previous section, the probability of having a specific configuration of duplexes $\boldsymbol{n}$ is:

$$p(\boldsymbol{n}) = \frac{1}{Z} f(\boldsymbol{n}) \prod_{i=1}^{S} \prod_{j=i}^{S} \prod_{k=1}^{K_{ij}} \left(\frac{\zeta_{ij}^{(k)}}{N} \right)^{n_{ij}^{(k)}}, \tag{3.9}$$

where, compared to Eq.3.3, the product depending on k-th mode of binding has been added. Analogously the partition function is:

$$Z = \sum_{\boldsymbol{n}} f(\boldsymbol{n}) \prod_{i=1}^{S} \prod_{j=i}^{S} \prod_{k=1}^{K_{ij}} \left(\frac{\zeta_{ij}^{(k)}}{N} \right)^{n_{ij}^{(k)}}. \tag{3.10}$$

where the sum is over all possible values of $n_{ij}^{(k)}$ satisfying the condition $0 \leq 2\sum_{k=1}^{K_{ii}} n_{ii} + \sum_{j \neq i} \sum_{k=1}^{K_{ij}} n_{ij} \leq N_i, \forall i$. The degeneracy $f(\boldsymbol{n})$ in Eq. 3.7 changes accordingly as:

$$f(\boldsymbol{n}) = \prod_{i=1}^{S} \frac{N_i!}{\left(N_i - 2\sum_{k=1}^{K_{ii}} n_{ii}^{(k)} - \sum_{j \neq i} \sum_{k=1}^{K_{ij}} n_{ij}^{(k)}\right)!} \times \frac{1}{2^{\sum_{k=1}^{K_{ii}} n_{ii}^{(k)}}} \prod_{j=i}^{S} \prod_{k=1}^{K_{ij}} \frac{1}{n_{ij}^{(k)}!}.$$

Finally the average number of duplexes formed by i-th and j-th according to k-th mode can be calculated from the partition function as:

$$\langle n_{ij}^{(k)} \rangle = \zeta_{ij}^{(k)} \frac{\partial \ln Z}{\partial \zeta_{ij}^{(k)}}. \tag{3.11}$$

3.3.3 Ensemble melting curve

The fraction of paired oligomer, i.e. the ensemble melting curve θ_e, is the sum of two times the average number of all duplexes $\langle n_{ij}^{(k)} \rangle$:

$$\theta_e = \frac{2\sum_{i=1}^{S} \sum_{j=i}^{S} \sum_{k=1}^{K_{ij}} \left\langle n_{ij}^{(k)} \right\rangle}{N}, \tag{3.12}$$

The melting curve θ_e can also be written as a linear combination of melting curves of each species i-th, $\theta_e^{(i)}$, as:

$$\theta_e = \sum_{i=1}^{S} \chi_i \theta_e^{(i)}, \tag{3.13}$$

with

$$\theta_e^{(i)} = \frac{\sum_{j=1}^{S} \sum_{k=1}^{K_{ij}} (1 + \delta_{ij}) \left\langle n_{ij}^{(k)} \right\rangle}{N_i}, \tag{3.14}$$

and δ_{ij} being the Kronecker's delta.
Since different oligomers species are highly and uniformly populated, as described in section 3.5.1, thermodynamic limit could be applied to Eq. 3.12. Although, $f(\boldsymbol{n})$ in Eq. 3.7 cannot be factorized, making impossible a derivation of θ_e in the thermodynamic limit, two distinct approximations allow to factorize $f(\boldsymbol{n})$ and consequently calculate the thermodynamic limit of θ_e. Let's assume that pairing between oligomers occurs only through a mutual alignment of bases. So, the different modes of pairing, k, can be described in terms of the integer shift parameter α_s, such that $-(L-1) \leq \alpha_s \leq L-1$ (see the second DNA duplex in Fig.3.1(c)). For example, $\alpha_s = 0$ describes the kth mode with perfect alignment between oligomers. The sign of α_s indicates if the overhang occurs on the 3' terminal ($\alpha_s > 0$) or on the 5' terminal ($\alpha_s < 0$). Consequently all the k-th modes of pairing between i-th and j-th oligomer are $K_{ij} = 2L - 1$.

Low T approximation

Let's firstly consider rsDNA solution at low T, where it is likely that most energetic pairings are formed. In such context, let's assume that most duplexes have $\alpha_s = 0$ and duplexes not perfectly aligned, i.e. duplexes with $\alpha_s \neq 0$, are negligible. Let's define $j^*(i)$ as the j-th sequence complementary to the i-th sequence. For self complementary oligomers $j^*(i) = i$ and all ij duplexes with $j \neq j^*(i)$ are negligible. So in Eq. 3.7, the factorial of free oligomers written in terms of α_s, $N_i - 2\sum_{\alpha_s=-L+1}^{\alpha_s=L-1} n_{ii}^{(\alpha_s)} - \sum_{j\neq i}\sum_{\alpha_s=-L+1}^{\alpha_s=L-1} n_{ij}^{(\alpha_s)}$, becomes $N_i - \sum_{j^*(i)}(1+\delta_{ij^*(i)})n_{ij^*(i)}^{(0)}$. To understand how such approximation affects the θ_e, let's consider a system composed by three sequences, labelled 1, 2 and 3, with 1 and 2 complementary oligomers and 3 self-complementary oligomer. So $\boldsymbol{n} = (n_{12}, n_{33})$ and $f(\boldsymbol{n})$ is:

$$\begin{aligned} f(\boldsymbol{n}) = &\prod_{i=1}^{3} \frac{N_i!}{(N_i - \sum_{j^*(i)} n_{ij*(i)}(1+\delta_{ij*(i)}))!} \times \frac{1}{2^{n_{ii}^{(0)}}} \prod_{j=i}^{3} \frac{1}{n_{ij^*(i)}!} = \\ = &\frac{N_1!}{(N_1 - n_{12})! n_{12}!} \times \frac{N_2!}{(N_2 - n_{12})!} \times \frac{N_3!}{(N_3 - 2n_{33})! 2^{n_{33}} n_{33}!}. \end{aligned} \tag{3.15}$$

$f(\boldsymbol{n})$ has been factorized in three terms. The product of the first two terms is the degeneracy for a complementary couple (CC), whereas the third term is the degeneracy for a self-complementary couple (SC), as described in [8].

So the partition function Z:

$$Z = \sum_{\boldsymbol{n}} f(\boldsymbol{n}) \prod_{i=1}^{3}\prod_{j=i}^{3}\left(\frac{\zeta_{ij}^{(0)}}{N}\right)^{n_{ij^*(i)}^{(0)}} = \sum_{n_{12}} \frac{N_1!N_2!}{(N_1-n_{12})!(N_2-n_{12})!n_{12}!}\left(\frac{\zeta_{12}^{(0)}}{N}\right)^{n_{12}^{(0)}} \times \sum_{n_{33}} \frac{N_3!}{(N_3-2n_{33})!2^{n_{33}}n_{33}!}\left(\frac{\zeta_{33}^{(0)}}{N}\right)^{n_{33}^{(0)}}, \tag{3.16}$$

where the sum over $\boldsymbol{n}$ has been written as the product of sums over n_{12} and n_{33} with corresponding degeneracy and Boltzmann factors for complementary and self-complementary oligomers. So, it's possible to calculate n_{12} and n_{33} with Eq. 3.11 and the melting curve θ_e with Eq. 3.12:

$$\theta_e = \frac{2\sum_{i=1}^{3}\sum_{j=i}^{3}\langle n_{ij}^{(0)}\rangle}{N} = \frac{2(\langle n_{12}\rangle + \langle n_{33}\rangle)}{N} = \theta_e^{12} + \theta_2^{33}, \tag{3.17}$$

which highlights that θ_e is given by the sum of the melting curve of a couple of complementary sequences, θ_e^{CC} (θ_e^{12}), and the melting curve of a self complementary sequence, θ_e^{SC} (θ_e^{33}). By extending this example to all perfect pairings, the melting curve for a rsDNA mixture in the *Low T approximation* is given by the sum of all perfect duplexes, composed by both complementary and self complementary sequences.
θ_e^{CC} and θ_e^{SC} can be analytically calculated in the thermodynamic limit as shown in [8], and they can be written in the unified equation as:

$$\theta_e^{(i)} = 1 - \frac{2}{1+\sqrt{1+\frac{4}{4^L}\zeta_{ij^*(i)}^{(0)}}}, \tag{3.18}$$

where in the Boltzmann factor the concentration c is replaced by $c/2$ for CC oligomers, whereas for SC oligomers is c. Therefore, θ_e can be calculated according to Eq.3.13.

High T approximation

Let's now consider the limit of high temperatures. In such a regime the fraction of paired oligomers is negligible and no relevant correlation among duplexes are expected, i.e. $N_i - 2\sum_{k=1}^{K_{ii}} n_{ii}^{(k)} - \sum_{j\neq i}\sum_{k=1}^{K_{ij}} n_{ij}^{(k)} \approx N_i$ and Eq.3.7 can be factorized. Each oligomer is in contact with an infinite bath of free sequences. Consequently, the fraction of free a oligomer i, 1 - $\theta_e^{(i)}$ with concentration $1/4^L$, can be calculated as the ratio between the probability of being a free oligomer and the sum of probabilities of being paired with any j oligomer. Such probabilities are expressed by the corresponding Boltzmann

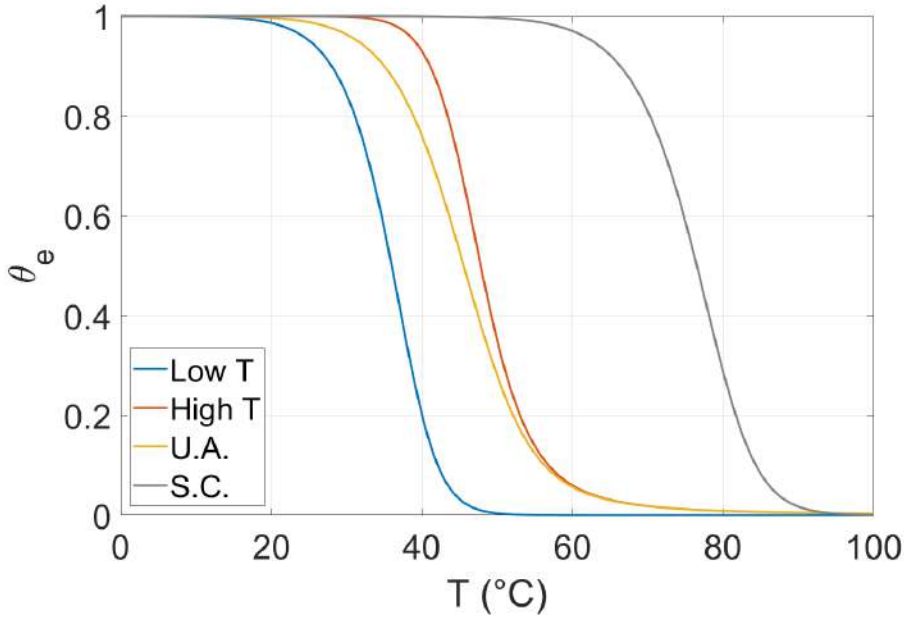

Figure 3.7: Melting Curves of rsDNA, according to the *Low T Approximation* (blue line), *High T Approximation* (red line) and *Unification Ansatz* (yellow line); the rsDNA behaviour can be compared with the melting curve of a Self-Complementary sequence in the same experimental conditions (grey line). The curves are computed for 12N, $c = 25$ g/l at NaCl 1M

factors. In particular, for a free oligomer with $\Delta G_{ij}{=}0\ \forall\ j$ its probability is 1 (see Eq.3.5). Therefore 1 - $\theta_e^{(i)}$ is:

$$1 - \theta_e^{(i)} = \frac{1}{1 + \frac{1}{4^L}\sum_{j=1}^{4^L}\sum_{\alpha_s=-L+1}^{L-1}\zeta_{ij}^{(\alpha_s)}}, \tag{3.19}$$

where the concentration of the i oligomer is the total concentration c scaled by 4^L, and the value 1 in the denominator comes from the contribution with $\alpha_s = L$, when no duplexes are formed ($1/4^L \sum_{j=1}^{4^L} = 1$). Consequently, the melting curve of species i is

$$\theta_e^{(i)} = 1 - \frac{1}{1 + \frac{1}{4^L}\sum_{j=1}^{4^L}\sum_{\alpha_s=-L+1}^{L-1}\zeta_{ij}^{(\alpha_s)}}. \tag{3.20}$$

Unification Ansatz

Both limits at low and high T can be unified with the following *Unification Ansatz*:

$$\theta_e^{(i)} = 1 - \frac{2}{1 + \sqrt{1 + \frac{4}{4^L}\sum_{j=1}^{4^L}\sum_{\alpha_s=-L+1}^{L-1}\zeta_{ij}^{(\alpha_s)}}}. \tag{3.21}$$

This equation is very symilar to the θ_e for self-complementary sequences (Eq.18a in [8]), with two dissimilarities: the effective concentration for any sequence is $c/4^L$; double sommation takes into account of all possible partner j and mutual alignments α_s. For high temperatures $\zeta_{ij}^{(\alpha_s)} << 1$ and expanding the square root, eq3.21 converges to 3.20. For low T, the only contribution comes from duplexes i-$j^*(i)$ with $\alpha_s = 0$. Consequently Eq.3.21 reduces to Eq.3.18. Fig.3.7 show the comparison of the three different approximation for 12N rsDNA solution at c_{rsDNA}=25g/l and c_{NaCl}=1M. The melting curve described by *Unification Ansatz* approach takes into account of all the variety of duplexes, not only duplexes with $\alpha_s = 0$, and underestimates the fraction of duplexes compared to the *High T approximation*.

3.3.4 Ensemble melting with f_{CG} content

Eq.3.21 requires the knowledge of all $(2L\text{-}1)\times 4^L \times 4^L$ binding energies, which is unfeasible. To tackle this problem, let's parametrize the duplex with following pairing errors vector:

$$\boldsymbol{\alpha} \equiv (\alpha_s, \alpha_{e1}, \alpha_{e2}, \alpha_i). \tag{3.22}$$

Besides α_s, α_{e1} and α_{e1} count the number of external mismatches at both duplex terminals, α_i counts the inner duplex mismatches, as sketched in Fig.3.1. Such a parametrization allows to calculate the degeneracy g(L,$\boldsymbol{\alpha}$), i.e. the number of oligomers of length L that form duplexes described by same $\boldsymbol{\alpha}$ with a particular reference oligomer:

$$g(L, \boldsymbol{\alpha}) = 4^{|\alpha_s|} 3^{(\alpha_{e1}+\alpha_{e2}+\alpha_i)} \binom{L - 2 - |\alpha_s| - \alpha_{e1} - \alpha_{e2}}{\alpha_i}. \tag{3.23}$$

Each overhang bases can be one out of 4 different bases ($4^{|\alpha_s|}$). Each internal and external mismatches is given by one out of three different choices of bases ($3^{(\alpha_{e1}+\alpha_{e2}+\alpha_i)}$). Binomial coefficient counts all the possible ways to place the internal mismatches in a sequence with the condition $\alpha_i \leq L - 2 - |\alpha_s| - \alpha_{e1} - \alpha_{e2}$ to guarantee at least one perfect base pairing in the duplex. Such degeneracy allows us to perform the coarse grained modelling of free energies according to the energy parametrization thoroughly described in the following section 3.3.5 in terms of the fraction f_{CG}, defined as the number of bases C and/or G, n_{CG}, divided by L. For a given reference sequence i with f_{CG}, the coarse grain model consists on associating a unique energy $\Delta G_{f_{CG}}^{(\boldsymbol{\alpha})}$, to all the duplexes that share the same $\boldsymbol{\alpha}$. The energy parametrization is based on the Nearest Neighbour (NN) approach [9, 4], where binding entropy and enthalpy are calculated as a sum of quartets neighbouring nucleobases forming the duplex. For example, let's consider the simplest case of perfect duplexes ($\boldsymbol{\alpha} = \mathbf{0}$). According to Eq. 3.21 and NN we should calculate 4^L different free

energies corresponding to the 4^L different duplexes in rsDNA mixture. The energy parametrization allows us to calculate a unique average free energy weighted on f_{CG} content, reducing the calculation of different free energies from 4^L to 1. Fig.3.8 shows the effect of such approximation. Dashed purple line is the average melting curve of 40 melting curves calculated with standard Santalucia protocol corresponding to 40 different DNA perfect duplexes ($\boldsymbol{\alpha} = \mathbf{0}$) with f_{CG}=0.5 and relative standard deviation (shaded area). Whereas, the green line is the melting curve of a perfect duplex with f_{CG}=0.5, whose energy is computed with energy parametrization. So, the temperature discrepancy, ΔT between the purple and green curves is $\Delta T < 1°$C.
By introducing this coarse grain model in Eq. 3.21, the melting curve $\theta_e^{(i)}$ can be written as:

$$\theta_e^{(f_{CG})} = 1 - \frac{2}{1 + \sqrt{1 + \frac{4}{4^L} \sum_{\boldsymbol{\alpha}} g(L, \boldsymbol{\alpha}) \zeta_{f_{CG}}^{(\boldsymbol{\alpha})}}}, \tag{3.24}$$

where the product $g(L, \boldsymbol{\alpha})\zeta_{f_{CG}}^{(\boldsymbol{\alpha})}$ is the statistical weight of the pairings of an oligomer of f_{CG} content with all the sequences that form duplex with quality $\boldsymbol{\alpha}$. This term is inside the $\sum_{\boldsymbol{\alpha}}$ to include all the possible pairs described by corresponding $\boldsymbol{\alpha}$. Fig. 3.8 displays the $\theta_e^{(f_{CG})}$ with f_{CG} ranging from 0 to 1 in the case of 8N. The ensemble melting, θ_e, is calculated as the average of $\theta_e^{(f_{CG})}$ over all possible values of f_{CG}, weighted with the probability $p_L(f_{CG})$ of having a sequence with f_{CG} in the rsDNA mixture:

$$\begin{aligned} \theta_e &= \sum_{f_{CG}} p_L(f_{CG})\, \theta_e^{f_{(CG)}} = \\ &= 1 - \frac{1}{2^L} \sum_{f_{CG}} \binom{L}{f_{CG} \cdot L} \times \frac{2}{1 + \sqrt{1 + \frac{4}{4^L} \sum_{\boldsymbol{\alpha}} g(L, \boldsymbol{\alpha}) \zeta_{f_{CG}}^{(\boldsymbol{\alpha})}}}. \end{aligned} \tag{3.25}$$

The number of total different sequences in terms of n_{CG} bases for a sequence L, can be written as $2^{n_{CG}} \times 2^{L-n_{CG}} = 2^L$ The probability $p_L(f_{CG})$ is given by the ratio between the binomial coefficient that counts all different sequences with equal n_{CG} bases but located in different position along the sequence and the total number of different sequences 2^L. Fig. 3.8 shows θ_e (red line) for 8N. Surprisingly we can observe the same features of Fig.3.4: the T_m predicted for rsDNA is $\approx 30°$C lower than those of complementary strands at the same c_{rsDNA} and c_{NaCl} (dashed line); the T dependence of rsDNA θ_e is milder, confirming the variety of energies of various duplexes in rsDNA mixture.

3.3.5 Energetic Parametrization based on α

In this section, it is discussed the calculation of the free energy difference $\Delta G_{f_{CG}}^{(\boldsymbol{\alpha})}$ between a random DNA duplex and its random DNA oligomer, represented by vector $\boldsymbol{\alpha} = (\alpha_s, \alpha_{e1}, \alpha_{e2}, \alpha_i)$, with length of sequence, L and

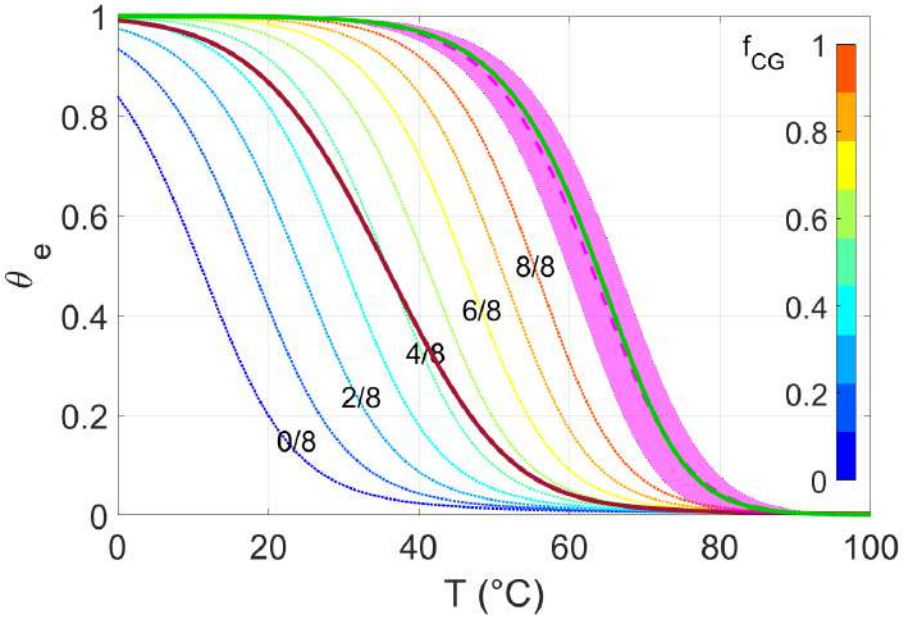

Figure 3.8: Theoretical predictions of melting curves, computed for DNA 8mers at $25g/L$ and 1M NaCl. The green solid line stands for the melting curve predicted for a pair of complementary strands when using the energy obtained from our parametrization, with $\boldsymbol{\alpha} = \mathbf{0}$ and $f_{CG} = 0.5$. The purple dashed-line and related shading are, respectively, the mean value and the standard deviation computed using the NN model over a set of 40 melting curves of distinct DNA sequences with fixed $f_{CG} = 0.5$. The dotted lines are the melting curve $\theta_e^{(f_{CG})}$ of 8N with different values of f_{CG} (Eq. 3.24), as specified in the colorbar. The solid dark red line is the ensemble melting curve θ_e, obtained by the average of $\theta_e^{(f_{CG})}$ over the possible values of f_{CG} (as given by Eq. 3.25).

fraction of C or G bases, f_{CG}.
It's important to split such energy in its enthalpic and an entropic contributions which also depends on $\boldsymbol{\alpha}$ and f_{CG}:

$$\Delta G_{f_{CG}}^{(\boldsymbol{\alpha})} = \Delta H^{(\boldsymbol{\alpha})}(L, f_{CG}) - T\Delta S^{(\boldsymbol{\alpha})}(L, f_{CG}) \tag{3.26}$$

$\Delta H^{(\boldsymbol{\alpha})}(L, f_{CG})$ and $\Delta S^{(\boldsymbol{\alpha})}(L, f_{CG})$ can be described as a sum of the following energies components: perfect pairing, internal terminal contributions, AT

terminal penalties and internal mismatches:

$$\begin{aligned}\Delta H^{(\boldsymbol{\alpha})}(L, f_{CG}) =& \Delta H_{\text{perf}}(L, f_{CG}) + \Delta H_{\text{term}}\left(\alpha_s, \alpha_{e1}, \alpha_{e2}\right) + \\ &+ \Delta H_{AT-pen} + \Delta H_{\text{int}}(\alpha_i), \end{aligned} \tag{3.27a}$$

$$\begin{aligned}\Delta S^{(\boldsymbol{\alpha})}\left(L, f_{CG}\right) =& \Delta S_{\text{perf}}(L, f_{CG}) + \Delta S_{\text{term}}\left(\alpha_s, \alpha_{e1}, \alpha_{e2}\right) + \\ &+ \Delta S_{AT-pen} + \Delta S_{\text{int}}(\alpha_i) + \Delta S_{\text{salt}}\left(L, \boldsymbol{\alpha}, [Na^+]\right) \end{aligned} \tag{3.27b}$$

The correction term due to salt concentration $[Na^+]$ is added in the entropic term. Each contribution will be now discussed in detail.

Perfect pairing

To simplify the following discussion, nucleobases are classified in two categories: weak bases $w = \{A, T\}$, and strong ones s $= \{C, G\}$. Averaging over all the possible combination of propagation sequences ww/ww, ss/ss, ws/ws along with sw/sw, reported in [4], it possible to calculate the following averaged energetic contributions: ΔH^{ww}_{NN} = -7.4 kcal/mol, ΔH^{ss}_{NN} = -9.1 kcal/mol, ΔH^{ws}_{NN} = -8.2 kcal/mol, ΔS^{ww}_{NN} = -21.07 cal/(molK), ΔS^{ss}_{NN} = -22.85 cal/(molK) and ΔS^{ws}_{NN}=-22.08 cal/(molK). We observe that $\Delta H^{ws}_{NN} \approx (\Delta H^{ww}_{NN} + \Delta H^{ss}_{NN})/2$ and $\Delta S^{ws}_{NN} \approx (\Delta S^{ww}_{NN} + \Delta S^{ss}_{NN})/2$. Consequently, the average enthalpy and entropy contribution of the quartets in a perfect paired duplex can be calculated as the linear combination of ΔH^{ss}_{NN} and ΔH^{ww}_{NN}, weighted on f_{CG}:

$$\Delta H_{NN}(f_{CG}) = \Delta H^{ss}_{NN} f_{CG} + \Delta H^{ww}_{NN}(1 - f_{CG}) \tag{3.28a}$$

$$\Delta S_{NN}(f_{CG}) = \Delta S^{ss}_{NN} f_{CG} + \Delta S^{ww}_{NN}(1 - f_{CG}) \tag{3.28b}$$

The contribution of a perfectly matched duplex is computed as:

$$\Delta H_{\text{perf}}(L, f_{CG}) = \Delta H_0 + (L-1)\Delta H_{NN}(f_{CG}), \tag{3.29a}$$

$$\Delta S_{\text{perf}}(L, f_{CG}) = \Delta S_0 + (L-1)\Delta S_{NN}(f_{CG}) \tag{3.29b}$$

where ΔH_0 and ΔS_0 are initial enthalpy and entropy contribution [4].

Terminal Contribution

Random DNA duplexes could end up with external mismatched base pairs and/or overhangs due to a shifted alignment of the random DNA oligomers. α_{e1} and α_{e2} counts the external mismatches in each end, whereas $|\alpha_s|$ counts the shifted bases. External mismatches can be computed as sum of two dangling bases at the same end [10]. So, the first step consists on calculating the averaged contributions of generic dangling bases $x = \{A, T, C, G\}$, next to

strong bases (xs/s and sx/s) or weak bases (xw/w and wx/w); the resulting averaged contributions are ΔH^s_{dang} = -3.99 kcal/mol, ΔH^w_{dang} = -1.08 kcal/mol, ΔS^s_{dang} = -11.15 cal/molK and ΔS^w_{dang} = -2.76 cal/molK. Consequently, the average energy contributions for a single dangling base weighted on f_{CG} is:

$$\Delta H_{dang}(f_{CG}) = \Delta H^s_{dang} f_{CG} + \Delta H^w_{dang}(1 - f_{CG}) \tag{3.30a}$$
$$\Delta S_{dang}(f_{CG}) = \Delta S^s_{dang} f_{CG} + \Delta S^w_{dang}(1 - f_{CG}) \tag{3.30b}$$

Finally, the contribution of terminal mismatches can be written as:

$$\begin{aligned}\Delta H_{\text{term}}(\alpha_s, \alpha_{e1}, \alpha_{e2}) =& c_t(\alpha_s, \alpha_{e1}, \alpha_{e2})\,\Delta H_{dang}(f_{CG}) + \\ &- (|a_s| + a_{e1} + a_{e2})\Delta H_{NN}(f_{CG})\end{aligned} \tag{3.31a}$$
$$\begin{aligned}\Delta S_{\text{term}}(\alpha_s, \alpha_{e1}, \alpha_{e2}) =& c_t(\alpha_s, \alpha_{e1}, \alpha_{e2})\,\Delta S_{dang}(f_{CG}) + \\ &- (|a_s| + a_{e1} + a_{e2})\Delta S_{NN}(f_{CG})\end{aligned} \tag{3.31b}$$

where for for each unpaired bases the corresponding averaged energetic contribution, $\Delta H_{NN}(f_{CG})$ or $\Delta S_{NN}(f_{CG})$, is subtracted . It's important to highlight that the condition $\alpha_{e1} + \alpha_{e2} + |\alpha_s| \leq L - 1$ must be satisfied to have at least one correct base pairing. The parameter $c_t(\alpha_s, \alpha_{e1}, \alpha_{e2})$ counts the dangling bases in the duplex:

$$c_t(\alpha_s, \alpha_{e1}, \alpha_{e2}) = \begin{cases} 0 & \text{if } \alpha_{e1} = \alpha_{e2} = \alpha_s = 0 \\ 2 & \text{if } \alpha_{e1} + \alpha_{e2} = 0 \text{ and } \alpha_s \neq 0 \\ 2 & \text{if } \alpha_{e1}\alpha_{e2} = 0 = \alpha_s \text{ and } \alpha_{e1} + \alpha_{e2} \neq 0 \\ 3 & \text{if } \alpha_{e1}\alpha_{e2} = 0 \text{ and } \alpha_{e1} + \alpha_{e2} \neq 0, \alpha_s \neq 0 \\ 4 & \text{if } \alpha_{e1}\alpha_{e2} > 0, \quad \forall \alpha_s. \end{cases} \tag{3.32}$$

Terminal AT Penalty

Another terminal correction involves the energy penalty due to AT terminal base pair. The corresponding penalty thermodynamic parameters are [4]: ΔH^{sing}_{AT-pen} = 2.2 kcal/mol, ΔS^{sing}_{AT-pen} = 6.9 cal/molK. So, the terminal AT penalty weighted on f_{CG} is:

$$\Delta H_{AT-pen}(f_{CG}) = \Delta H^{sing}_{AT-pen} c_{AT}(1 - f_{CG}) \tag{3.33a}$$
$$\Delta S_{AT-pen}(f_{CG}) = \Delta S^{sing}_{AT-pen} c_{AT}(1 - f_{CG}) \tag{3.33b}$$

where c_{AT} counts the AT penalties in the random DNA duplex:

$$c_{AT}(\alpha_s, \alpha_{e1}, \alpha_{e2}) = \begin{cases} 0 & \text{if } \alpha_s > 0 \\ 1 & \text{if } \alpha_{e1}\alpha_{e2} = 0 = \alpha_s \text{ and } \alpha_{e1} + \alpha_{e2} \neq 0 \\ 2 & \text{if } \alpha_s = \alpha_{e1} = \alpha_{e2} = 0 \end{cases} \tag{3.34}$$

Internal Mismatches Contribution

Internal mismatches random DNA duplex are here counted by α_i. The condition $|\alpha_s| + \alpha_{e1} + \alpha_{e2} + \alpha_i \leq L - 2$ must be satisfied by duplexes with one matched base pair at least. Averaging over all possible propagation sequences (ww/ww, ss/ss, sw/sw, ws/ws) with only one WC pairng [11, 12, 13, 14], it's possibile to calculate the averaged energetic contribution ΔH_{int}= 0.15 kcal/-mol and ΔS_{int}= -0.84 cal/(molK). Further internal mismatches are assumed to be never consecutive. Consequently, each internal mismatch corresponds to loose two quartets. Therefore, the contribution due to internal mismatches is:

$$\Delta H_{\text{int}}(\alpha_i) = 2\alpha_i[\Delta H_{int} - \Delta H_{NN}(f_{CG})] \tag{3.35a}$$

$$\Delta S_{\text{int}}(\alpha_i) = 2\alpha_i[\Delta S_{int} - \Delta S_{NN}(f_{CG})], \tag{3.35b}$$

removing the energetic contribution of 2 perfect quartets for each unpaired internal bases.
The arbitrary choice on internal mismatches placement should not affect the pairing statistics because the predicted fraction of duplexes with even a single internal mismatch is negligible (see Fig.3.9(a)).

Salt Contribution

The salt contribution ΔS_{salt} has been determined from empirical expression obtained by Owczarzy et al. [7]. Their T_m correction described by Eq. 22, is:

$$\frac{1}{T_m(2)} = \frac{1}{T_m(1)} + (4.29 f_{CG} - 3.95) \cdot 10^{-5} \ln[Na^+]_2 + +9.4 \cdot 10^{-6} \ln^2[Na^+]_2, \tag{3.36}$$

where $T_m(1)$ is the melting temperature at sodium reference concentration $[Na^+]_1 = 1\text{M}$, whereas $T_m(2)$ is the melting temperature at sodium ion concentration $[Na^+]_2$. The melting temperature for a 2 state hybridization model can be calculated as a function of the enthalpic and entropic parameters of the pairing (Eq. 19 in [8]):

$$\frac{1}{T_m(1)} = \frac{R \cdot ln(\gamma[c]) + \Delta S[1M]}{\Delta H}, \tag{3.37a}$$

$$\frac{1}{T_m(2)} = \frac{R \cdot ln(\gamma[c]) + \Delta S[1M] + \Delta S_{salt}([Na^+]_2)}{\Delta H}, \tag{3.37b}$$

where $[c]$ is the total strand concentration, γ is a multiplicative constant which is different for self-complementary or couple of complementary strands.

Combining Eq. (3.36), Eq. (3.37)a and Eq. (3.37)b, it is possible to isolate the salt contribution:

$$\begin{aligned}\Delta S_{salt}([Na^+]_2) =& \Delta H \cdot (4.29 f_{CG} - 3.95) \cdot 10^{-5} \ln[Na^+]_2 + \\ &+ 9.4 \cdot 10^{-6} \ln^2[Na^+]_2\end{aligned} \tag{3.38}$$

where ΔH can be calculated according to Eq. (3.27)a.

3.3.6 Probability of defectless and defected duplexes

The developed model allows us to have access to pairing statistics of duplexes. For example the fraction $\theta_{\boldsymbol{\alpha}}^{(f_{CG})}$ of oligomers that form duplexes described by any specific $\boldsymbol{\alpha}$ can be calculated as:

$$\theta_{\boldsymbol{\alpha}}^{(f_{CG})} = \frac{g(L,\boldsymbol{\alpha})\zeta_{f_{CG}}^{(\boldsymbol{\alpha})}}{\sum_{\boldsymbol{\alpha}'} g(L,\boldsymbol{\alpha}')\zeta_{f_{CG}}^{(\boldsymbol{\alpha}')}} \theta_e^{(f_{CG})}, \tag{3.39}$$

where the prefactor of the total fraction of paired oligomers $\theta_e^{(f_{CG})}$ is the statistical weight of specific duplexes family sharing the same $\boldsymbol{\alpha}$. Therefore, the fraction of duplexes with a specific $\boldsymbol{\alpha}$ in the rsDNA mixture is given by the sum on f_{CG} of the product between the probability $p_L(f_{CG})$ and its $\theta_{\boldsymbol{\alpha}}^{(f_{CG})}$:

$$\theta_{\boldsymbol{\alpha}} = \sum_{f_{CG}} p_L(f_{CG})\, \theta_{\boldsymbol{\alpha}}^{(f_{CG})} = \frac{1}{2^L} \sum_{f_{CG}} \binom{L}{f_{CG} \cdot L} \cdot \frac{g(L,\boldsymbol{\alpha})\zeta_{f_{CG}}^{(\boldsymbol{\alpha})}}{\sum_{\boldsymbol{\alpha}'} g(L,\boldsymbol{\alpha}')\zeta_{f_{CG}}^{(\boldsymbol{\alpha}')}} \theta_e^{(f_{CG})} \tag{3.40}$$

Fig.3.9(a) shows the six error that occur with highest probability for 12N rsDNA solution as a function of T. Surprisingly the perfect duplexes (red line) are not negligible over a wide range of T, and at low T they are beyond the 10%. Most duplexes have terminal errors (green and light blue lines). Indeed their penalty energetic cost is lower compared to internal pairing errors. Fig.3.9(b) shows the fraction $\theta_{|\boldsymbol{\alpha}|}$ for L = 8, 12 and 20, where $|\boldsymbol{\alpha}|$ counts the total unpaired bases: $|\boldsymbol{\alpha}| \equiv |\alpha_s| + \alpha_{e1} + \alpha_{e2} + \alpha_i$. The most frequent duplexes have $|\boldsymbol{\alpha}| = 1$ accordingly to panel (a) for 12N, whereas duplexes with $|\boldsymbol{\alpha}| =$ 3 and 4 are negligible at low T. Moreover, the low T regime highlights how the $\theta_{|\boldsymbol{\alpha}|}$ collapse on the same curve for all the three different length. Indeed, at low T duplexes with internal mismatches are negligible, as pointed out by panel(a), and consequently reasonably considering $\alpha_i = 0$, both the degeneracy $g(L,\boldsymbol{\alpha})$ (see Eq.3.23) and the ratio of Boltzmann factors turn to be independent on L. Let's see more explicitly the ratio of Boltzmann factors,

$\varphi_{\boldsymbol{\alpha}}$. Firstly, $\Delta G(L, \boldsymbol{\alpha})$ can be wirrten as:

$$\begin{aligned}\Delta G(L, \boldsymbol{\alpha}) =& \Delta G_0 + (L - 1 - |\alpha_s| - \alpha_{e1} - \alpha_{e2})\Delta G_{NN} + c_t(\alpha_s, \alpha_{e1}, \alpha_{e2})\Delta G_{dang} + \\ &+ \Delta G_{AT-pen} = \\ &= L \cdot \Delta G_{NN} + \Delta\widetilde{G}(\alpha_s, \alpha_{e1}, \alpha_{e2}),\end{aligned} \tag{3.41}$$

where ΔG_{NN}, ΔG_{dang}, ΔG_{AT-pen} are calculated according to Eq.3.28, 3.30b, 3.33b. The second term in the second equality gather all the terms not dependent on L. So, $\varphi_{\boldsymbol{\alpha}}$ can be written as:

$$\begin{aligned}\varphi_{\boldsymbol{\alpha}} &= \frac{g(\boldsymbol{\alpha})[c]e^{-\beta[L\Delta G_{NN} + \Delta\widetilde{G}\alpha_s, \alpha_{e1}, \alpha_{e2})]}}{\sum_{\boldsymbol{\alpha'}} g(\boldsymbol{\alpha'})[c]e^{-\beta[L\Delta G_{NN} + \Delta\widetilde{G}(\alpha'_s, \alpha'_{e1}, \alpha'_{e2})]}} = \\ &= \frac{\cancel{e^{-\beta L\Delta G_{NN}}} g(\boldsymbol{\alpha})e^{-\beta\Delta\widetilde{G}(\alpha_s, \alpha_{e1}, \alpha_{e2})}}{\cancel{e^{-\beta L\Delta G_{NN}}} \sum_{\boldsymbol{\alpha'}} g(\boldsymbol{\alpha'})e^{-\beta\Delta\widetilde{G}(\alpha'_s, \alpha'_{e1}, \alpha'_{e2})}},\end{aligned} \tag{3.42}$$

where the exponential L-dependent factor in the denominator has been factored out from the sum because it does not depend on $\boldsymbol{\alpha'}$. The only term related to L is in the $\sum_{\boldsymbol{\alpha'}}$, that is performed with the constrain $|\alpha_s| + \alpha_{e1} + \alpha_{e2} \leq L - 1$. However, as shown in Fig.3.9 the only significant duplexes are those with $|\boldsymbol{\alpha}| << 1$, so that also the sum can be considered independent on L.

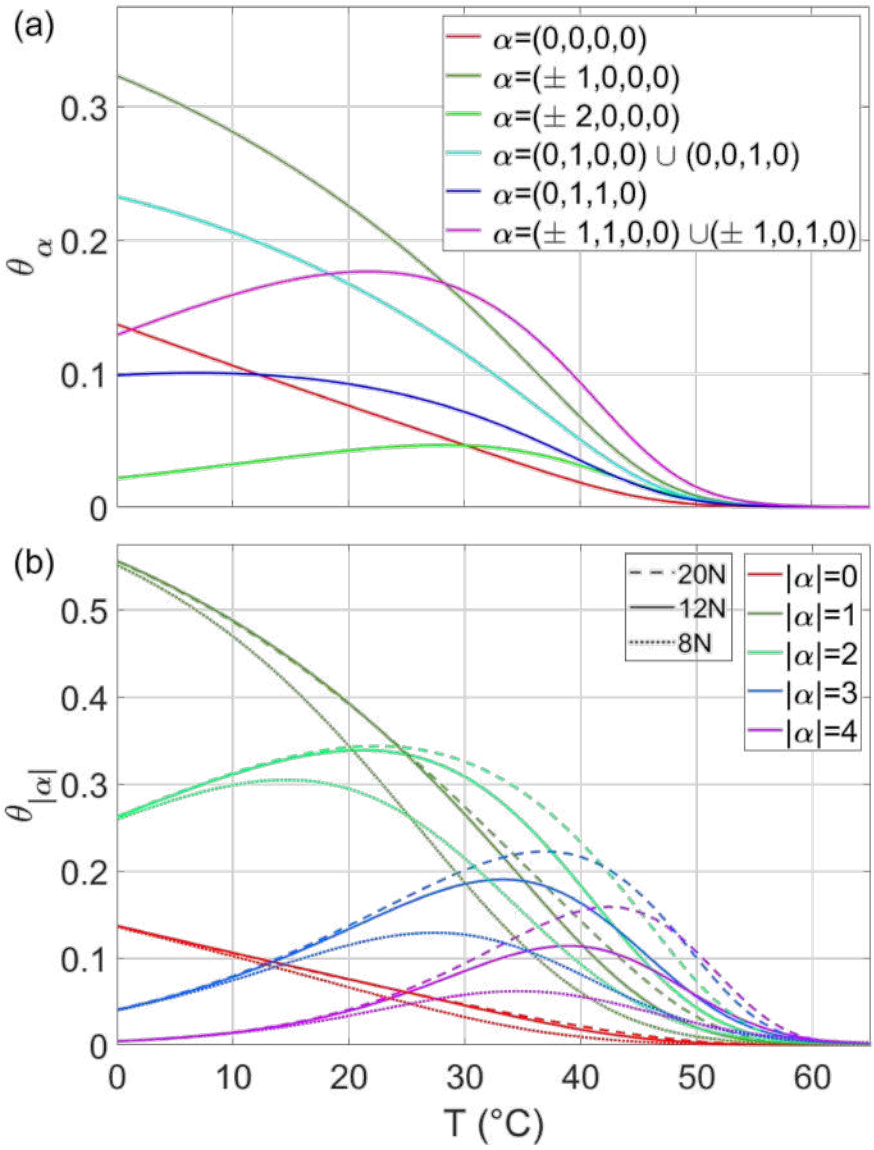

Figure 3.9: Theoretical predictions of the fraction of perfect and defected rsDNA duplexes (as given by Eq. (3.40)), parametrized by $\boldsymbol{\alpha}$. (a): the six most probable duplex motifs in 12N. (b): fraction of duplexes with a total number of unpaired bases $|\boldsymbol{\alpha}|$, in 8N (dotted lines), 12N (continuous lines), 20N (dashed lines). $c_{NaCl} = 1M$ and $c_{rsDNA} = 25g/l$.

3.3.7 Perfect pairing probability at different energies

How does the change of pairing energies affect the pairing statistics of perfect duplexes? The theoretical model allows us to answer this question, whose

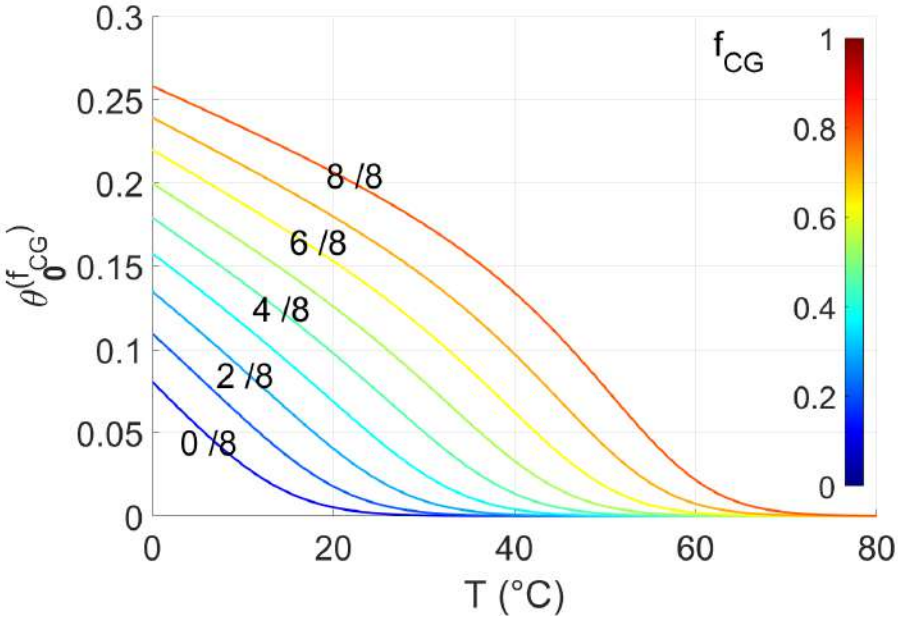

Figure 3.10: $\theta_{\mathbf{0}}^{(f_{CG})}$ computed with different values of f_{CG} in 8N, at a concentration of $25g/l$ and NaCl 1M.

result will be further discussed in the conclusions paragraph. Let's calculate the fraction of perfect duplexes $\theta_{\mathbf{0}}^{(f_{CG})}$ at different f_{CG}. The result is shown in Fig. 3.10. As expected the less energetic perfect duplexes, formed by all AT pairings (f_{CG}=0), melt at lower temperature compared to duplexes with higher f_{CG} content, with a difference of ≈ 40°C between $\theta_{\mathbf{0}}^{(f_{CG}=0)}$ and $\theta_{\mathbf{0}}^{(f_{CG}=1)}$. Focusing the attention on $\theta_{\mathbf{0}}^{(f_{CG})}$ at $T = 0$°C, the fraction of perfect duplexes increases from 8% to 26% at higher and higher f_{CG} content. So, even in the most energetic case, the fraction of perfect duplexes is not the dominant one and there's a 10% difference compared to data shown in Fig.3.9.

3.4 Comparison between Experiments and Theory

The model developed in the previous section allows a direct comparison with experimental observations reported in section3.2 for melting experiments. Fig.3.11 shows the experimental ensemble melting curves (symbols) with the corresponding theoretical predictions computed with Eq.3.25 (dashed lines). By highlighting that dashed lines have no free parameters, the agreement between all experiments and theory is remarkably good. The corresponding melting temperatures for both experiments and theory are plotted in Fig. 3.12 as a function of c_{NaCl}. The difference between experiments and prediction

can be quantified by the absolute average difference between experimental, T_m^{exp}, and predicted, T_m^{pred}, melting temperatures [7]:

$$\langle \Delta T_m \rangle_{AVE} = \sum_{i=1}^{M} \frac{|T_{m,i}^{pred} - T_{m,i}^{exp}|}{M} \tag{3.43}$$

where M is the total number of melting temperature, which is 9 for data shown in Fig. 3.11. $\langle \Delta T_m \rangle_{AVE} = 1°C$ or $\langle \Delta T_m \rangle_{AVE} = 0.5°C$ excluding the lowest T_m corresponding to 12N at c_{NaCl}=0.15M, which low-T baseline is hardly recognizable in a very limited T interval, and thus resulting in a higher uncertainty as discussed in section 3.5.4. In both cases, values of $\langle \Delta T_m \rangle_{AVE}$ are comparable with typical errors in prediction T_m for any DNA duplexes calculated with typical thermodynamic models [7].
The ensemble melting θ_e and specially its T behaviour reveal the huge variety and complexity of interaction that occur in rsDNA mixture and that are surprisingly well described by the developed theoretical model. Such an agreement support the theoretical result shown in Fig. 3.9 that give a microscopic insight on paring statistics, predicting the fraction of duplexes with quality $\boldsymbol{\alpha}$ as given by Eq.3.40. At low T, most duplexes have less than 2 pairing errorr ($|\alpha| \leq 2$) and approximately th 14% of duplexes are defectless. Such last result is further confirmed by contact quenching experiment described in [15], where the experimental fraction of perfect duplexes (which oligomers are fluorescent labelled) in rsDNA mixture for 8N is nicely in agreement with the theoretical predictions.
The compensation between the Boltzmann factor, which depends exponentially on binding free energy ΔG, and the degeneracy, g(L, $\boldsymbol{\alpha}$) (Eq.3.23), set the rule for pairing statistics between rsDNA oligomers. Both these factors have an exponential behaviour with two opposite trends as a function of $\boldsymbol{\alpha}$. The higher the number of pairing errors, the higher the degeneracy and the lower the Boltzmann factor (ΔG less negative). The smooth T and $\boldsymbol{\alpha}$ dependence of $\theta_{\boldsymbol{\alpha}}$ reveal that these two factors compensate with energy that partially prevail on degeneracy.

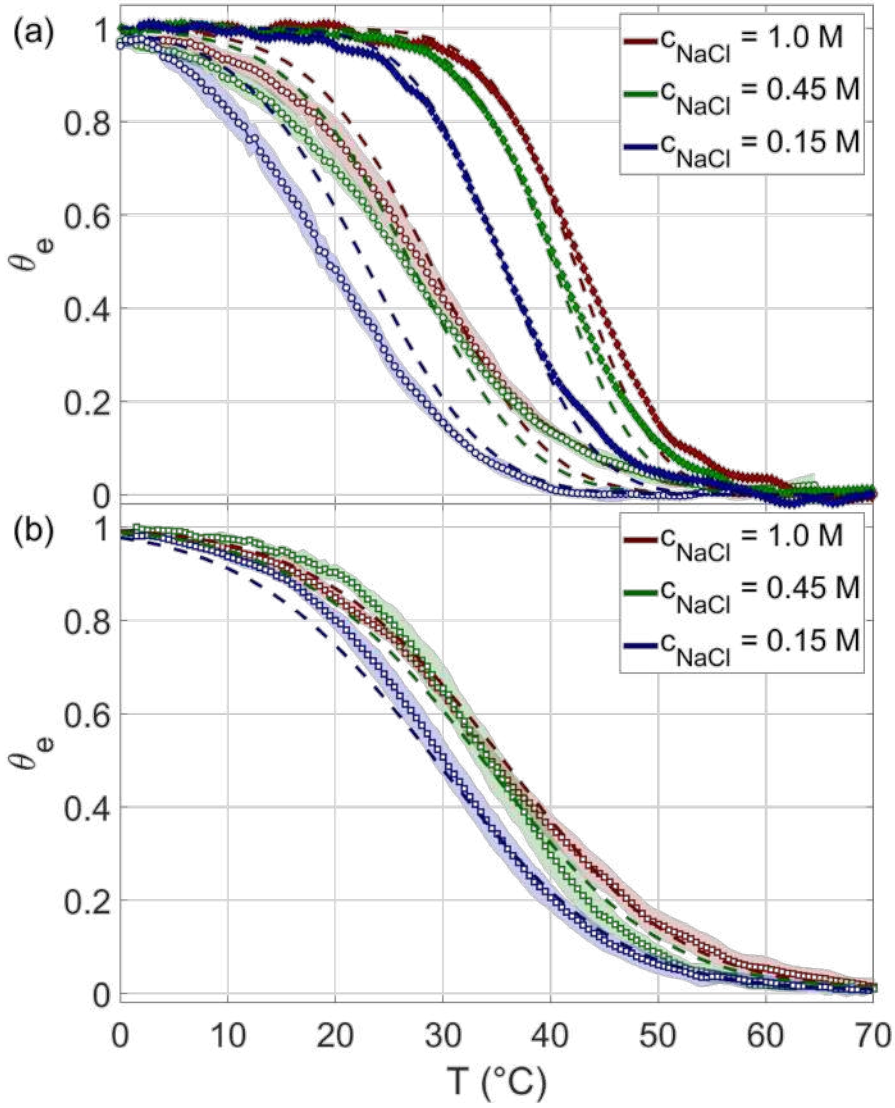

Figure 3.11: Comparison between experimental melting curves shown in Fig 3.5 and corresponding θ_e(T) predicted according to Eq. 3.25

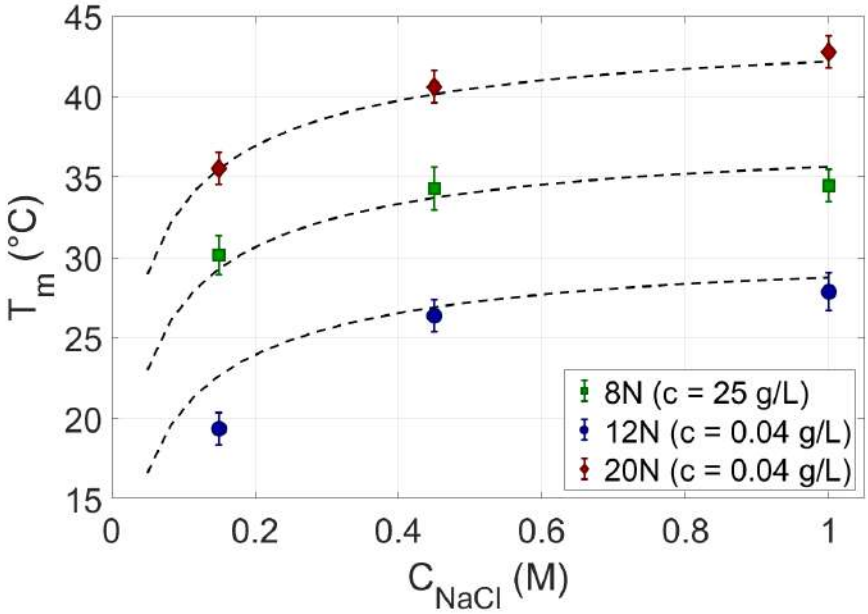

Figure 3.12: Comparison between experimental T_m shown in Fig 3.6 and corresponding theoretical predictions (dashed lines) calculated according to Eq. 3.25 at $\theta_e(T = T_M) = 0.5$. 8N (green squares), 12N (blue circles) and 20N (red diamonds).

3.4.1 Evidence of out-of-equilibrium conditions

Absorbance data for 20N rsDNA solutions at low concentration revealed out-of-equlibrium features. Before running melting experiments, rsDNA solutions were kept for several days at T=4°C. Fig.3.13 shows absorbance data for 20N at $c_{rsDNA} = 0.04$ g/l and $c_{NaCl} = 0.45M$. All heating and cooling ramps were performed at a ramp rate equals to $1°/min$. It's evident the difference between the first heating ramp (blue circles) and consecutive cooling and heating ramps. In particular, the smaller absorbance values of first ramp and the higher T_m compared to the consecutive ramps ones are a clear signature of a better quality of pairing between rsDNA oligomers. Indeed, the higher the average hypocromicity, the smaller the extinction molar coefficient and consecutively the UV absorbance A (Eq.3.44). The less the defects in rsDNA pairing, the higher the ΔG of the duplex and the T_m, too. The inset shows the slow recovery of UV Absorbance (green dots) of rsDNA solution quenched at T=35°C from T=80°C. The time needed to reach the A(T=35°C) of the first ramp is longer than 48 hours. Such a behaviour was observed also at c_{NaCl} =0.15 and $1M$. The quenching temperature was chosen slightly below the T_m to accelerate the formation of duplexes with high ΔG. In future works it will be investigated how different quenching temperature affects the kinetics and

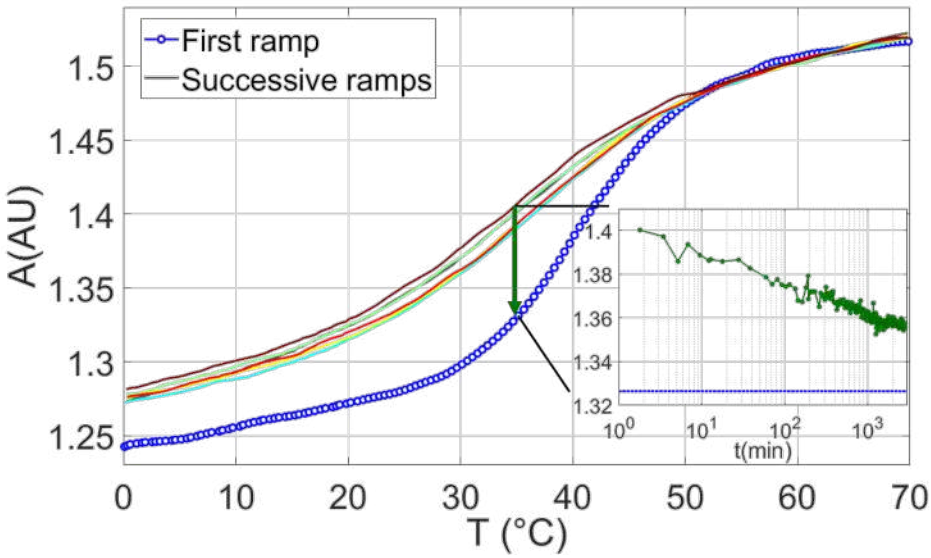

Figure 3.13: Evidence of out of equilibrium for 20N at c_{rsDNA} = 0.04 g/l and c_{NaCl} = 450mM. First heat ramp (blue dots) after one month equilibration time at $4°C$. Successive heat and cool ramps (colored lines) clearly shows a remarkably difference in both amplitude and melting temperature compared to first ramp. Inset: recovery absorbance profile in time (green dots) of the sample quenched at T = $35°C$ from T = $80°C$. A(T = $35°C$) of first ramp (dashed blue line) needs more than 48 hours to be recovered.

statistics of rsDNA pairings. Out-of-equilibrium features were also revealed in melting experiments at high concentration for 12N, which are different from 20N at low concentration. Fig. 3.14, at c_{rsDNA} =25 g/l and c_{NaCl} =0.15M, shows that the averaged heating (red circles) and cooling (blues circles) ramps collapse on two slightly different melting curves, whose T_m are 1°C far apart. Such a difference is even more evident at higher c_{NaCl}. Fig. 3.15 shows that the averaged heating and cooling θ_e are $\approx$ 4°C far apart. Increasing the c_{NaCl} up to 1M absorbance data, A, of cooling ramps don't show any inflection point and are almost linear, making impossible to fit them. It means that rsDNA oligomers didn't have enough time to form enough stable duplexes detectable by UV absorbance. For this reason in Fig.3.16, the average heating ramp is only shown. So, all these experiments suggest that rsDNA duplexes formed at low T are different and with higher thermal stability than those forming during cooling experiments. Moreover, the corresponding T_m are $\approx$15°C lower than the predicted ones. Such observations show that also rsDNA solution with $L \geq 12$ at high concentration didn't attain equilibrium, forming rsDNA duplexes less energetic than predicted ones. Moreover, their pairing statistics is strongly affected by thermal history, whose dependence is enhanced at

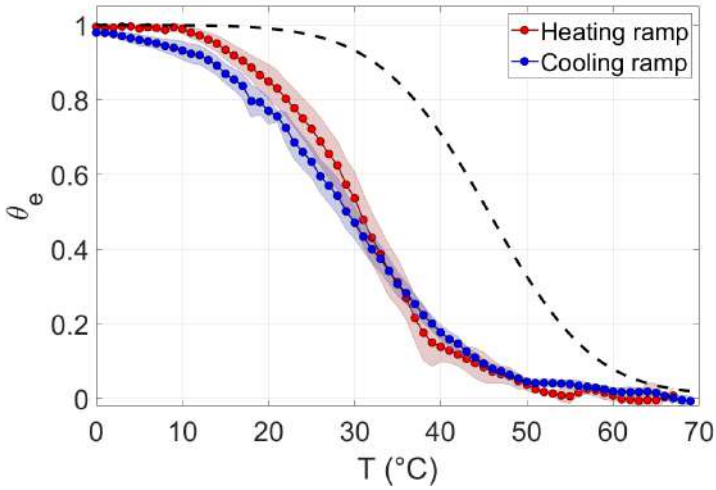

Figure 3.14: Evidence of out-of-equilibrium for 12N rsDNA at c_{12N} = 25 g/l and c_{NaCl} = 150 mM. Average heat ramp (red dots) and cool ramp (blue dots) are $4°C$ far apart in the range of T [20-40]$°C$ each other and their melting temperature are $10-14°C$ apart from the theoretical prediction (dashed line). Shaded region marks experimental uncertainty.

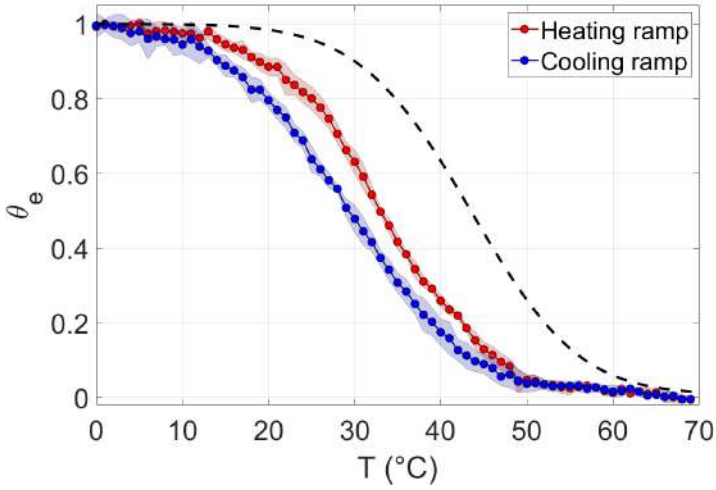

Figure 3.15: Evidence of out-of-equilibrium for 12N rsDNA at c_{12N} = 25 g/l and c_{NaCl} = 450 mM. Average heat ramp (red dots) and cool ramp (blue dots) are $4°C$ far apart in the range of T [20-40]$°C$ each other and their melting temperature are $10-14°C$ apart from the theoretical prediction (dashed line). Shaded region marks experimental uncertainty.

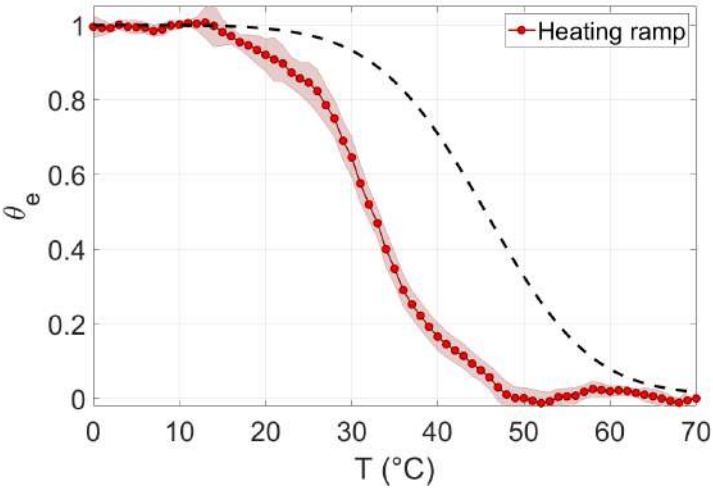

Figure 3.16: Evidence of out-of-equilibrium for 12N rsDNA at c_{12N} = 25 g/l and c_{NaCl} = 1M. Average heat ramp (red dots) whose melting temperature is 15°C apart from the theoretical prediction (dashed line). Cooling ramp is not reported as discussed in main text. Shaded region marks experimental uncertainty.

increasing salt concentration.

3.5 Matherial and Methods

3.5.1 Characterization of rsDNA synthesis

rsDNA oligomers were characterized by HPLC experiments. Fig. 3.17(a) shows HPLC traces of 12N compared with single strand oligomer 12ss (CTATGCCACCTA) and self-complementary 12-mer DD (CGCGAATTCGCG). The Main HPLC peaks of DD and 12ss occurs in the braoder profile of 12N and their relative distance is larger than their width. The width of 12N HPLC traces is the signature of rsDNA heterogeneity eluted at slightly different times due to both different A,T,C and G contents and realtive order in the sequence. Main peak of both 12 ss, which has $f_{CG} = 0.5$, and 12N occurs at the same time. In Fig. S 3.17(b) we show HPLC traces of rsDNA oligomers with L = 8,12 and 20. All HPLC traces share the same shape, and as expetected, the higher the the length, the longer the time at which the center of distribution occurs.
All DNA solution were prepared at $c_{DNA} = 0.1g/l$, and HPLC experiments were performed at $T = 40°C$. Moreover, the 20N was previously characterized by MALDI-TOF mass spectrometry (see [16]).
Provided that there are not significant bias in the rsDNA preparation, the

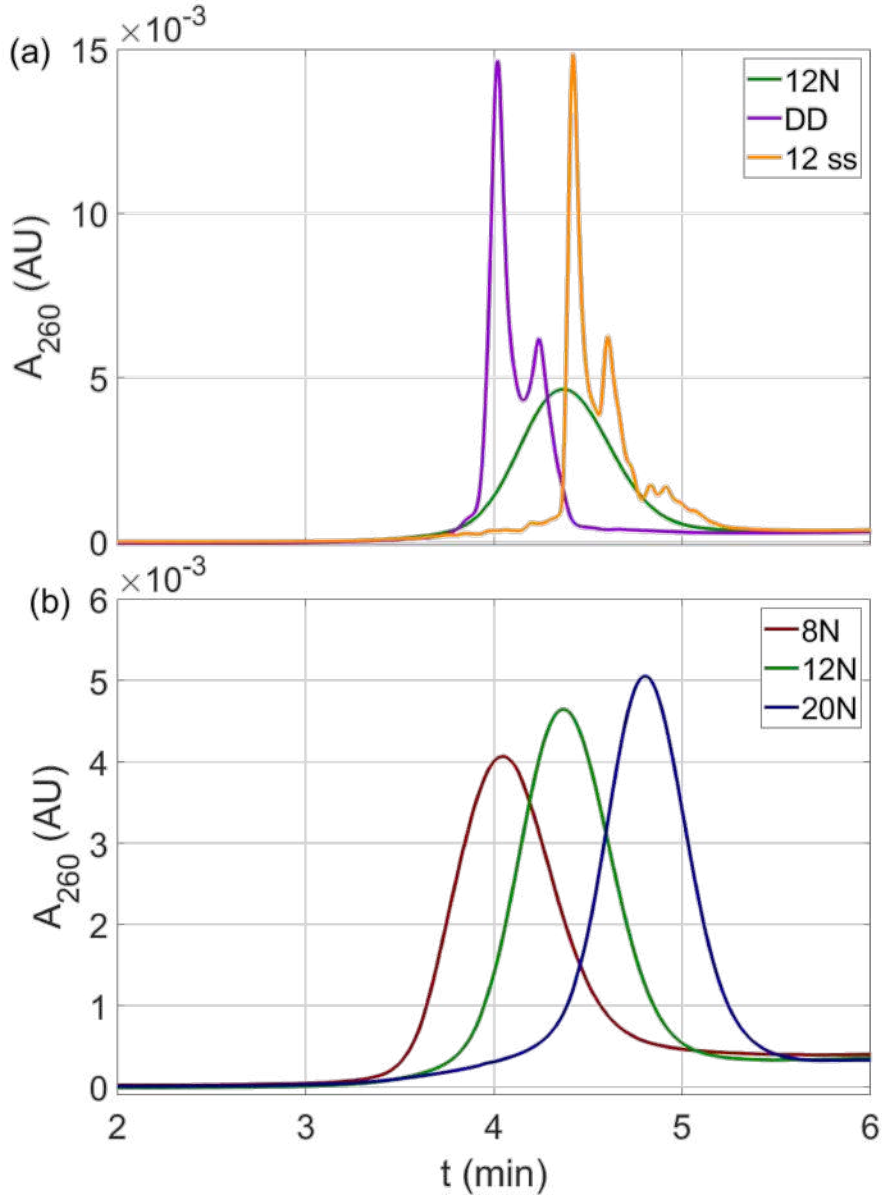

Figure 3.17: (a) HPLC traces of 12N compared with two different 12mers, 12ss (CTATGCCACCTA) and DD(CGCGAATTCGCG). (b) HPLC traces of 8N, 12N and 20N. HPLC experiments were performed at $T = 40°C$ and DNA solutions were prepared at c_{DNA}= 0.1g/l.

total number of DNA molecules N is way larger than the number of distinct rsDNA sequences. Indeed, let us consider the extreme case of 20N in diluted condition, i.e. the rsDNA system with the largest number of sequences $S = 4^{20} \approx 1.1 \cdot 10^{12}$ and lowest number of total DNA molecules N compared

to 8N and 12N at the same concentration. In melting experiments with the $1\,cm$ couvette, the sample occupies a volume $V = 1\,mL$ at a concentration of $c = 0,04g/L \approx 6mM$ and consequently $N = cV = 6\mu\,mol \approx 4 \cdot 10^{16}$. So, in rsDNA solution, each rsDNA oligomer has $N/S \approx 10^4$ replicas per molecular type.

3.5.2 Measurement of rsDNA concentration

The absorbance A of a DNA solution is given by the Beer–Lambert law:

$$A = \epsilon\, l\, c_{DNA} \tag{3.44}$$

ϵ is the molar extinction coefficient $(l/(molcm))$ of the DNA in the sample, where l is the optical path length in cm and c_{DNA} is the DNA molar concentration. The extinction coefficient for a particular DNA sequence i is computed, according to the NN model [17], with the following expression:

$$\epsilon_i = \epsilon_i(ext_1) + \epsilon_i(ext_2) + \sum_{q=1}^{L-1} \epsilon_i(q) \tag{3.45}$$

where $\epsilon_i(ext_1)$ and $\epsilon_i(ext_2)$ represents the contribution of the first and the last quartets of the DNA sequence and $\epsilon_i(q)$ are the extinction coefficients of the couples of bases in the position q and $q+1$ along the sequence i.

Here, we are interested in rsDNA solution where all possible distinct sequences of length L are equally populated. Consequently, the molar extinction coefficient for a generic rsDNA sequence of length L, $\epsilon_{rsDNA,L}$ is defined as the average of ϵ_i calculated over all the possible rsDNA sequences, $\langle\epsilon_i\rangle_i$:

$$\epsilon_{rsDNA,L} = \langle\epsilon_i(ext_1)\rangle_i + \langle\epsilon_i(ext_2)\rangle_i + \sum_{q=1}^{L-1}\langle\epsilon_i(q)\rangle_i = 2\bar{\epsilon}_{ext} + (L-1)\bar{\epsilon}_{int}, \tag{3.46}$$

where $\bar{\epsilon}_{ext} = \langle\epsilon_i(ext_1)\rangle_i = \langle\epsilon_i(ext_2)\rangle_i = 5375(cm\,M)^{-1}$ and the internal term is $\bar{\epsilon}_{int} = \langle\epsilon_i(q)\rangle_i = 9572(cm\,M)^{-1}$ [17]. In table 3.1 are reported the values for the rsDNA oligomers used in this work, calculated according to Eq.3.46. To convert concentration from mol/l to g/l, it is necessary to calculate the molecular weight. According to [18] the mean molecular weight of a nucleotide $N(=\{A,T,C,G\})$ incorporated in a nuclei acid is $\langle dN\rangle = 308.95$ g/mol. So the molecular weight has been calculated as:

$$MW_{rsDNA} = \langle dN\rangle L + 60 \tag{3.47}$$

where the addition of "60" is due to the hydroxyl terminus (on behalf of phosphate group). The values of MW for 8N, 12N and 20N are reported in Tab.3.1.

L	ϵ_{rsDNA} (l/(molcm))	MW (g/mol)
8	77754	2412
12	116042	3647
20	192618	6119

Table 3.1: Molar extinction coefficients and molecular weights for rsDNA oligomers with different length L, calculated according to Eq.3.46 and 3.47, respectively.

The concentration of rsDNA solution has been estimated via UV absorbance at high temperature, T>70°C, in order to melt any rsDNA duplexes, whose hypocromicity term is unknown because of the variety of pairing errors.
Stock aqueous solutions were prepared at crsDNA ≈ 50 g=l, whose concentration was measured diluting it at c ≈ 0.04 g/l. From stock solutions, the final samples concentrations were obtained by dilution, ranging from c_{rsDNA} = 0.04g/l to 25g/l.

3.5.3 UV Absorbance: Experimental Setup

Melting experiments have been performed with the Evolution 300 UV-Vis spectrophotometer from Thermo Scientific. In particular, its peltier allows to explore the range of temperature from 15° to 90°C (see Fig. 3.18). However, two crucial technological upgrades were necessary to get reliable $\theta_e(T)$ for 8N

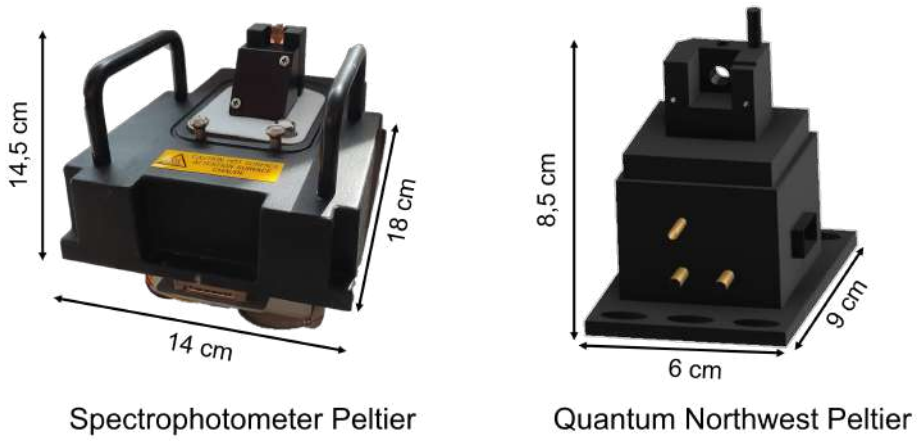

Figure 3.18: Comparison between the spectrophotometer peltier which works in the temperature range [15°C-90°C] and Quantum Northwest Peltier which works in the temperature range [-10°C-90°C].

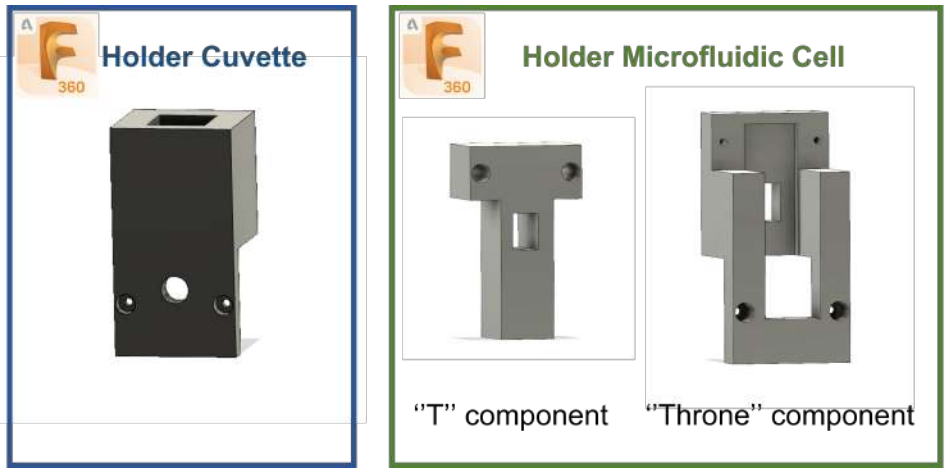

Figure 3.19: Aluminium holder components to host standard quartz cuvette (in blue box) and microfluidic cell (in green box) devised and designed with AutoDesk Fusion 360.

and 12N: 1) Quantum Northwest (QNW) peltier hot/cold stage (shown in Fig. 3.18) which allowed us to explore a wider range of temperature T = 0-90 °C; 2) the quartz microfluidic cell shown in Fig. 3.20 with path length l = 10 μm, which allowed us to increase c_{DNA} up to 20 - 30g/l. It's important to highlight that without having access to the lowest temperature, the analysis of most absorbance data A(T) would have been affected by high uncertainty as explained below. The new peltier system needs two different holders to host standard quartz cuvettes and microfluidic cell. Consequently, I devised and designed with AutoDesk and helped in manufacturing in a mechanic's shop the aluminium components shown in Fig. 3.19. The cuvette holder is composed by only one component that is screwed to the peltier and cuvette is inserted in the square hole (see Fig. 3.22). The circular hole allows the transmitted beam to pass through. Differently, the microfluidic cell holder is composed by two components, nicknamed "T" and "Throne" according to their shape. As shown in Fig.3.20, the "throne" component is screwed to the peltier and host the microfluidic cell in vertical position, and the "T" component is fastened into it. In such a way the thermal contact is optimized. The rectangular holes in both aluminium components allow the transmitted beam to pass through. In melting experiments, because of high temperature involved, evaporation must be prevented. In quartz cuvettes, such an achievement is guaranteed covering the DNA solution with mineral oil, which has lower density than water. Differently, in microfluidic cell, the two nozzles are sealed with EPDM corks (by S.A.P.A.T.) as shown in Fig. 3.20. The difference ΔA_{258} between absorbance values at $\lambda = 258nm$

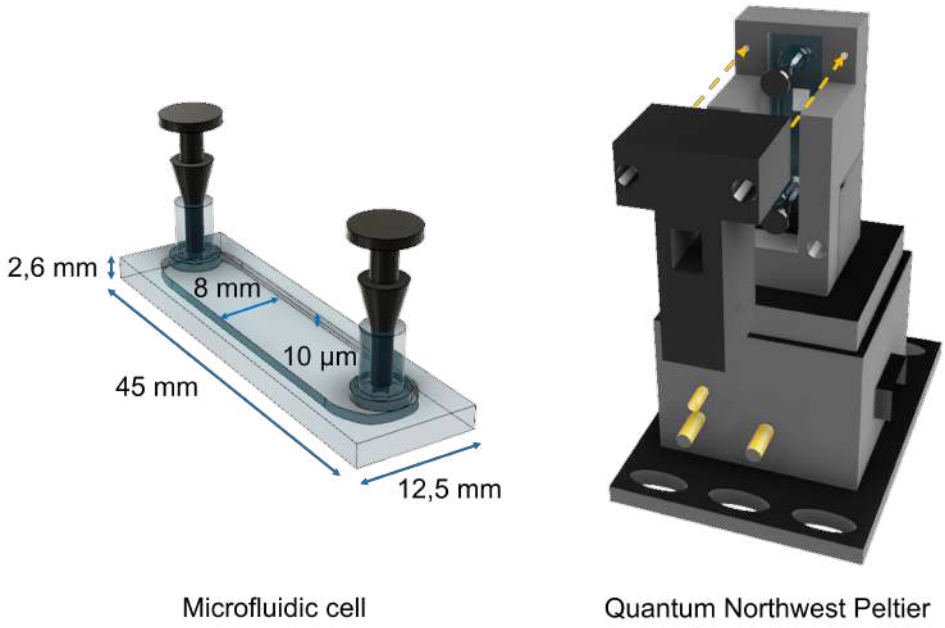

Figure 3.20: Quartz microfluidic cell with path length l = 10μm sealed with black EPDM corks. Quantum Northwest Peltier assembled with microfluidic holder. The "T" component is screwed to the "throne" component, where microfluidic cell is inserted in vertically, as indicated by yellow arrows.

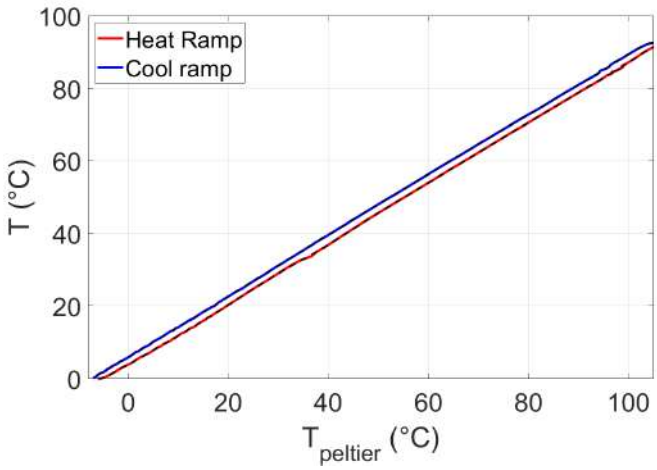

Figure 3.21: Temperature calibration of T measured by a thermistor in contact with the microfluidic cell vs. the internal control $T_{peltier}$. Red and blue lines: heating and cooling ramps, respectively. Dashed line: polynomial fit.

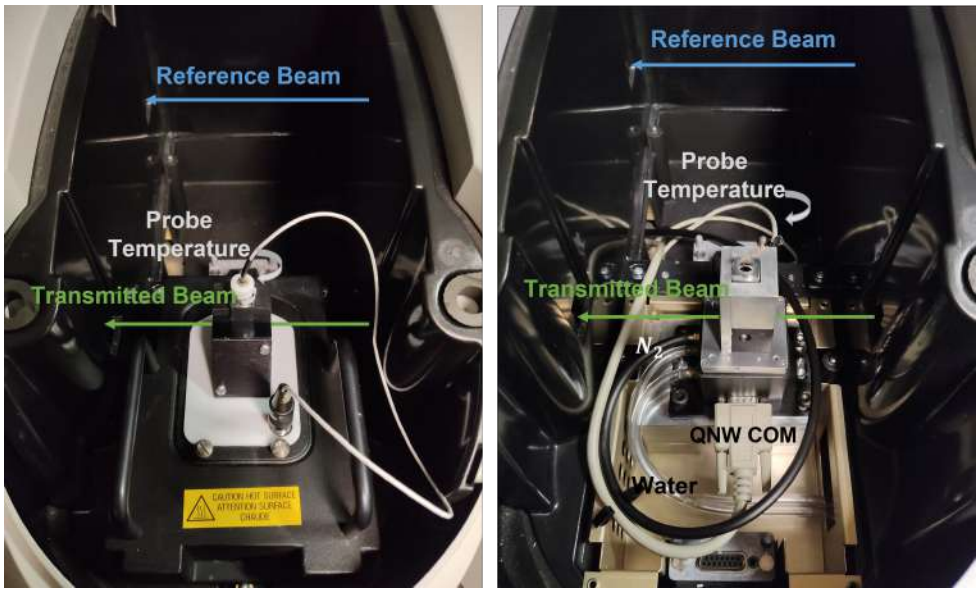

Figure 3.22: Experimental setup comparison. On the left spectrophotometer peltier with corresponding probe temperature. On the right Quantum Northwest peltier assembled with aluminium component for cuvette. Probe temperature is a thermistor plugged into Agilent A34970. Quantum Northwest COM is plugged into the PC to control the peltier. Transparent plastic tubes are plugged into a circulating bath that allows water to flows into. In black plastic tube N_2 flows to prevent condensation at low temperature.

before and after melting experiments in both diluted and high c_{DNA} condition was $\Delta A_{258} < 1\%$. The temperature is measured with an external thermistor plugged into Agilent A34970 data acquisition unit. In quartz cuvette experiments, such a probe temperature is directly inserted in the DNA solution, below water-oil interface. In cell experiments it's not possible to set the probe into the cell. Consequently, it has been performed the following temperature calibration: the thermistor, has been placed in the rectangular hole of the "T" component, and fixed on the microfluidic cell. It has been run consecutively a heating and cooling ramp at the commonly used ramp rate ($1°C/min$) from -6$°C$ to 100$°C$ and viceversa. During the ramps it has been measured the temperature T by thermistor as a function of the inner peltier probe temperature, $T_{peltier}$, as shown in Fig. 3.21. Because the whole system shown in Fig. 3.20 is thermally isolated, the temperature of solution inside cell is in thermal equilibrium with microfluidic device. Dissipation of heat was made possible by circulating water in plastic tubes plugged into both QNW peltier and circulating bath, and condensation at low T was prevented by nitrogen constant flux into the QNW peltier. Fig. 3.22 shows the comparison between the experimental setup of spectrophotometer with its peltier

and with QNW peltier, assembled with cuvette holder, and all the features described so far.

Software

The new setup is independent on the spectrophotometer and so cannot be controlled by its commercial software. Consequently, I have programmed a new application with App Matlab Designer which allow to control the spectrophotometer, the QNW peltier and the Agilent Data Acquisition simultaneously. Each device is connected to the computer with a communication port. The first step consists on defining such a serial port inside Matlab environment making the communication between device and Matlab possible. The second step consists on writing in Matlab the functions that allow to manage the devices. For QNW peltier and Agilent these functions can be found in [19] and [20], respectively. For example, the command function to ask the Current Temperature of peltier is: [F1 CT ?] and the response is: [F1 CT 22.84], which means that the current temperature is 22.84°C. The character encoding for both QNW Peltier and Agilent is ASCII. Although for the spectrophotometer there wasn't any documentation available, the Device Monitoring Studio application allowed to sniff the commands to control the spectrophotometer. An example is reported in Fig.3.23: commands and responses are written in Hexadecimal numeral system on the left with corresponding conversion in ASCII on the right. For example the function *SCA 220.0 350.0 0 1 240 user.5* run an absorbance measurement SCAnning from 220 to 350 nm with data interval of 1 nm and scan speed 240nm/min.

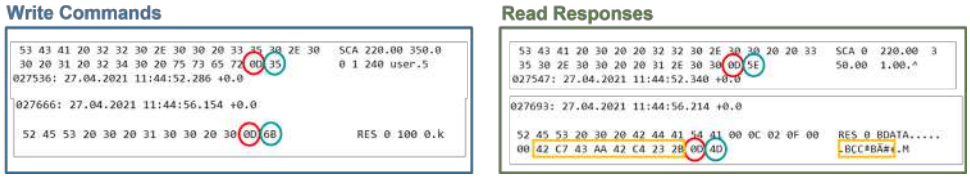

Figure 3.23: Spectrophotomer commands sniffed by Device Monitoring Studio application

It's important to understand the meaning of characters in coloured circles to write command functions rightly. The recursive combination of characters *0D* in red circles is the return statement and it must be added at the end of each function. The characters in blue circles change for every function and they represent the checksum command. Checksum commands are generated by specific algorithm, that it has been decoded after several attempts:

- convert all hexadecimal numbers until *0D* excluded, in decimal numbers.

- calculate the total sum
- divide the sum by 64
- add 48
- convert the decimal value in hexadecimal

It seems magic (and impossible on decoding), but a similar procedure is used to calculate the last symbol in fiscal codes. It's now possible to end up correctly each function. Another crucial and difficult decoding was related to the absorbance data. The command function *RES 0 100* make the detector measure absorbance, which values are hided in the highlighted yellow box in Fig.3.23. The hexadecimal characters string was analyzed with the online hexadecimal converter [21], that transforms such a string in decimal numbers according to different algorithms. The correct one is the **Float-Big Endian**, that provides numbers which correspond to the transmittance data in percentage, $\%T$. Absorbance data are finally obtained by A = -log10($\%T$).
The third step consists on assembling all these functions in an application that allows to measure absorbance during an arbitrary number of heating and cooling ramps (see Fig 3.24). Absorbance values at each temperature are calculated as the average of three absorbance around absorption peak.

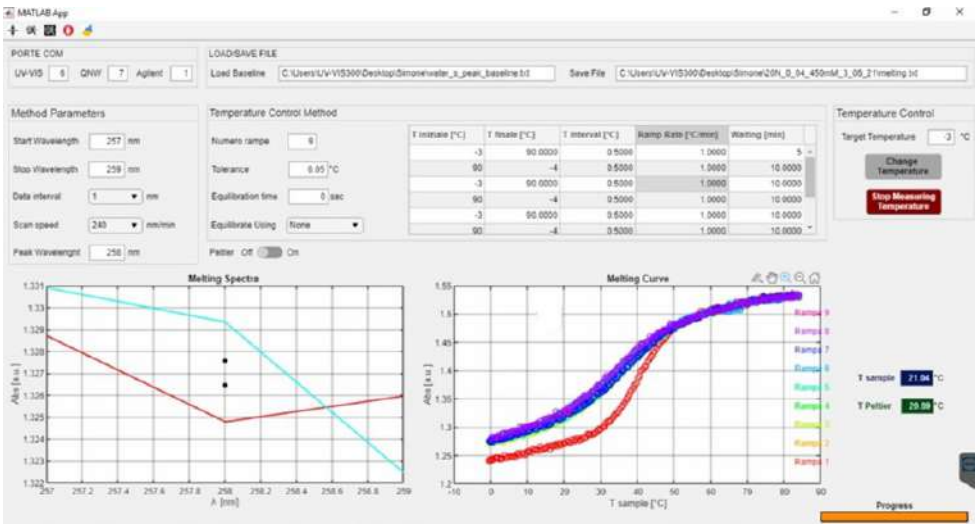

Figure 3.24: Home made software designed with AppDesigner in Matlab to control simultaneously QNW peltier and UV-VIS 300 spectrophotometer and perform melting experiments.

3.5.4 Melting Curve Data Analysis

Absorbance data are firstly smoothed with Savitzky-Golay (sgolayfilt Matlab function) filter with polynomial order 3. Fig. 3.25(a) displays A(T) smoothed curves of several consecutive heating and cooling ramps carried out on on the same rsDNA solution. The overlap of A(T) data make clear the reproducibility of experiments and that no evaporation occurs all over experimental duration. Successively, each A(T) curve is fitted in conformity with the usual data analysis [22] to get rid of the contribution of the absorbance drift at high temperature $A_{H.T.}(T)$ and at low temperature $A_{L.T.}(T)$:

$$\theta(T) = \frac{A(T) - A_{L.T.}(T)}{A_{H.T.}(T) - A_{L.T.}(T)}, \tag{3.48}$$

where θ is the fraction of free rsDNA single strands. To accomplish the fit of A(T) curves, it is necessary to hypothesize a functional shape for θ, and I opt for a sigmoid function described by:

$$\theta(T) = \frac{a_1}{1 + e^{-a_2(T-a_3)}} \tag{3.49}$$

and $A_{H.T.}$, $A_{L.T.}$ are:

$$\begin{aligned} A_{L.T.}(T) &= a_4 + a_5 T \\ A_{H.T.}(T) &= a_6 + a_7 T. \end{aligned} \tag{3.50}$$

The experimental absorbance data A(T) are consequently fitted by:

$$y(T) = \frac{a_1}{1 + e^{-a_2(T-a_3)}}(a_6 - a_4 + (a_7 - a_5)T) + a_4 + a_5 T, \tag{3.51}$$

where a_i, with $i = \{1, 7\}$, are free parameters. Fig. 3.25(b) displays one of the experimental A(T) curves with the corresponding fitting dashed line (Eq.3.51) and $A_{H.T.}$ and $A_{L.T.}$ as straight blue lines. Normalization of A(T) curves according to their fits allow to get the corresponding θ curves. Consequently, the fraction of oligomers in paired rsDNA duplexes θ_e is simply calculated as:

$$\theta_e(T) = 1 - \theta(T) \tag{3.52}$$

Fig. 3.25c displays such θ_e melting curves with no drifts. The average of these θ_e is shown in Fig3.25d as blue dots with corresponding experimental uncertainty marked by blue shading. The melting temperature T_m of the measurement is set by the condition $\theta_e(T = T_m) = 1/2$. Fig 3.25b shows the melting of duplexes starts before $15°C$, unveiling how crucial is to gain access to the range of temperature 0-$15°C$. Without the experimental setup described in previous section, the analysis of such absorbance data would have been highly affected by uncertainty.

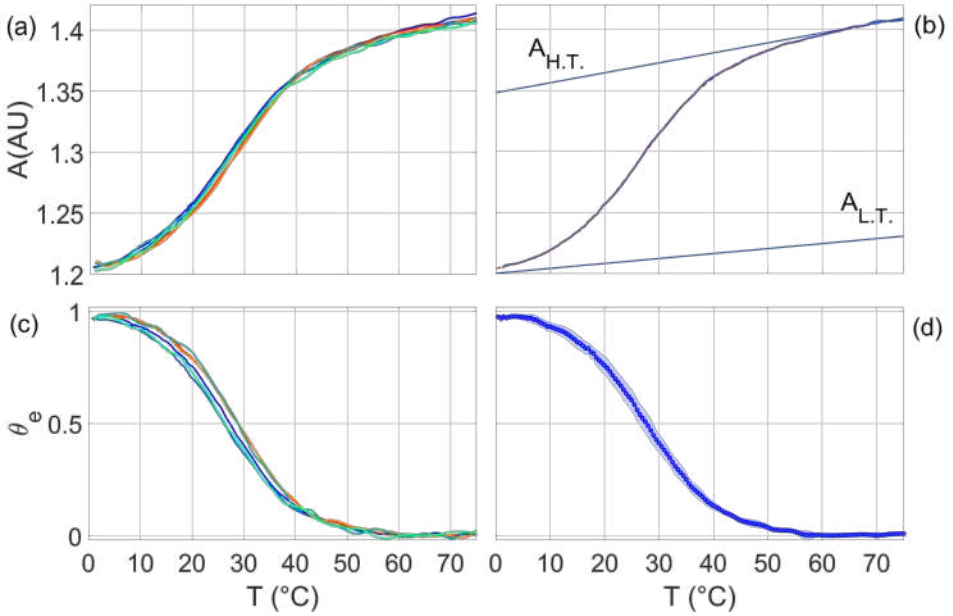

Figure 3.25: Step in the analysis of absorbance data $A(T)$ of 12N ($c_{rsDNA} = 0.04$ g/l, $c_{NaCl} = 1$M) used to extract the melting curves θ_e. (a) $A(T)$ measured in consecutive heating and cooling ramps, after smoothing. (b) $A(T)$ (blue line) fitted with Eq. 3.51 (dashed red line). Blue lines: linear fit at the high and low T drift as in Eq. 3.50. (c) θ_e curves obtained after drift subtraction as in Eq. 3.48. (d) θ_e obtained as an average over the curves in panel (c). Experimental uncertainty is shown as the shaded light blue area around the data point.

3.6 Conclusions

In this chapter rsDNA solution has been introduced as an example of superdiverse environment where a huge variety of interactions take place and lead to not trivial pair formations. The enormous heterogeneity of random DNA sequences, given by the combination of just four different nucleobases, their well energetically characterized pairing rules, the solid-state synthesis, and all experimental and theoretical tools explained in this chapter made possible the investigation of such system as paradigm of superdiverse mixture.
The agreement between experiments and theoretical model reveal that the selectivity of pairing lead to the formation of stable structures. Indeed, most duplexes have just one or two pairing errors and there is a not negligible fraction of perfect duplexes. Along with the stability of such predictions upon changing energies involved in nucloeobase paring (see Fig. 3.10), these results set a reference for the selectivity occurring in conditions of unavoidable randomness and variety of oligomers as those in prebiotic scenarios [23, 24].

Whatever the modifications in base paring, the process leading to chain amplification and increase in length, such study successfully describe the selectivity in oligomers pairing. Liquid crystal ordering is one of possible mechanism that enhance the length of nucleic acids [25, 26]. Indeed, rsDNA oligomers surprisingly shows liquid crystal (LC) ordering as reported in [27] for L=20, but even for shorter length L=12 as it will be shown in chapter 6. As reported in [28], if on one hand LC ordering is enhanced by blunt end duplexes and by duplexes with 1-2 complementary overhangs, on the other hand it is suppressed by both external mismatches and not complementary overhangs. How these different duplexes, which are the dominant ones, interact to give rise to LC ordering is a very fascinating and stimulating topic to be investigated in future.
Another important topic that deserves to be investigated more deeply are the evidences of out-of-equilibrium for rsDNA at both low and high concentration. It would be interesting to find out the strategies that lead such solution in equilibrium and to investigate how the pairing statics changes.
It's important to highlight that this work lead to the development of the a new setup that allows to measure melting of nucleic acids at high concentration ($25-30$g/l), that are sufficiently high for some nucleic acids, as guanosine monophosphate, to support LC ordering.
The powerful developed theoretical model, beyond its capability to predict the observed experimental behaviour for rsDNA mixture, consists on its applicability for any system formed by oligomers which pairing energy can be parametrized. For example, this model can be easily applied to mixtures with duplexes formed also by artificial nucleobases as Hachimoji DNA and RNA[29].

Bibliography

[1] Ron Milo and Rob Phillips. *Cell biology by the numbers*. Garland Science, 2015.

[2] R John Ellis. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Current opinion in structural biology*, 11(1):114–119, 2001.

[3] Anastasia Politou and Piero Andrea Temussi. Revisiting a dogma: the effect of volume exclusion in molecular crowding. *Current opinion in structural biology*, 30:1–6, 2015.

[4] John SantaLucia Jr and Donald Hicks. The thermodynamics of dna structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440, 2004.

[5] Joseph N Zadeh, Conrad D Steenberg, Justin S Bois, Brian R Wolfe, Marshall B Pierce, Asif R Khan, Robert M Dirks, and Niles A Pierce. Nupack: analysis and design of nucleic acid systems. *Journal of computational chemistry*, 32(1):170–173, 2011.

[6] Richard Owczarzy, Peter M Vallone, Frank J Gallo, Teodoro M Paner, Michael J Lane, and Albert S Benight. Predicting sequence-dependent melting stability of short duplex dna oligomers. *Biopolymers: Original Research on Biomolecules*, 44(3):217–239, 1997.

[7] Richard Owczarzy, Yong You, Bernardo G Moreira, Jeffrey A Manthey, Lingyan Huang, Mark A Behlke, and Joseph A Walder. Effects of sodium ions on dna duplex oligomers: improved predictions of melting temperatures. *Biochemistry*, 43(12):3537–3554, 2004.

[8] Carlos A. Plata, Stefano Marni, Amos Maritan, Tommaso Bellini, and Samir Suweis. Statistical physics of dna hybridization. *Phys. Rev. E*, 103:042503, Apr 2021. doi: 10.1103/PhysRevE.103.042503. URL https://link.aps.org/doi/10.1103/PhysRevE.103.042503.

[9] Saptarshi Ghosh, Shuntaro Takahashi, Tamaki Endoh, Hisae Tateishi-Karimata, Soumitra Hazra, and Naoki Sugimoto. Validation of the

nearest-neighbor model for Watson–Crick self-complementary DNA duplexes in molecular crowding condition. *Nucleic Acids Research*, 47(7): 3284–3294, 02 2019. ISSN 0305-1048. doi: 10.1093/nar/gkz071. URL https://doi.org/10.1093/nar/gkz071.

[10] Nupack. http://www.nupack.org/partition/new.

[11] Hatim T Allawi and John SantaLucia. Nearest-neighbor thermodynamics of internal a-c mismatches in dna: Sequence dependence and ph effects. *Biochemistry*, 37(26):9435–9444, 1998.

[12] Hatim T Allawi and John SantaLucia Jr. Thermodynamics of internal c-t mismatches in dna. *Nucleic acids research*, 26(11):2694–2701, 1998.

[13] Hatim T Allawi and John SantaLucia. Nearest neighbor thermodynamic parameters for internal g-a mismatches in dna. *Biochemistry*, 37(8): 2170–2179, 1998.

[14] Hatim T Allawi and John SantaLucia. Thermodynamics and nmr of internal g-t mismatches in dna. *Biochemistry*, 36(34):10581–10594, 1997.

[15] S. Di Leo, S Marni, CA Plata, TP Fraccia, GP Smith, A. Maritan, S. Suweis, and Bellini T. Pairing statistics and melting of random dna oligomers: Fi
nding your partner in superdiverse environments. *PLOS Computational Biology*, 2022.

[16] Tommaso Bellini, Giuliano Zanchetta, Tommaso P Fraccia, Roberto Cerbino, Ethan Tsai, Gregory P Smith, Mark J Moran, David M Walba, and Noel A Clark. Liquid crystal self-assembly of random-sequence dna oligomers. *Proceedings of the National Academy of Sciences*, 109(4): 1110–1115, 2012.

[17] Andrey V. Tataurov, Yong You, and Richard Owczarzy. Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. *Biophysical Chemistry*, 133:66–70, 2008.

[18] Molbiotools' user guide, dna calculator. https://www.molbiotools.com/manual/dnacalcman.html.

[19] Qnw tc1 serial commands. https://www.qnw.com/wp-content/uploads/2019/01/Text-commands.pdf.

[20] Keysight 34970a command reference. https://documentation.help/Keysight-34970A-34972A/documentation.pdf.

[21] Scadacore online hex converter. https://www.scadacore.com/tools/programming-calculators/online-hex-converter/.

[22] Richard Owczarzy. Melting temperatures of nucleic acids: discrepancies in analysis. *Biophysical chemistry*, 117(3):207–215, 2005.

[23] Gerald F Joyce. The antiquity of rna-based evolution. *nature*, 418(6894): 214–221, 2002.

[24] Walter Gilbert. Origin of life: The rna world. *nature*, 319(6055):618–618, 1986.

[25] Tommaso P Fraccia, Gregory P Smith, Giuliano Zanchetta, Elvezia Paraboschi, Youngwooo Yi, David M Walba, Giorgio Dieci, Noel A Clark, and Tommaso Bellini. Abiotic ligation of dna oligomers templated by their liquid crystal ordering. *Nature communications*, 6(1):1–8, 2015.

[26] Marco Todisco, Tommaso P Fraccia, Greg P Smith, Andrea Corno, Lucas Bethge, Sven Klussmann, Elvezia M Paraboschi, Rosanna Asselta, Diego Colombo, Giuliano Zanchetta, et al. Nonenzymatic polymerization into long linear rna templated by liquid crystal self-assembly. *ACS nano*, 12 (10):9750–9762, 2018.

[27] Tommaso Bellini, Giuliano Zanchetta, Tommaso P Fraccia, Roberto Cerbino, Ethan Tsai, Gregory P Smith, Mark J Moran, David M Walba, and Noel A Clark. Liquid crystal self-assembly of random-sequence dna oligomers. *Proceedings of the National Academy of Sciences*, 109(4): 1110–1115, 2012.

[28] Simone Di Leo, Marco Todisco, Tommaso Bellini, and Tommaso P Fraccia. Phase separations, liquid crystal ordering and molecular partitioning in mixtures of peg and dna oligomers. *Liquid Crystals*, 45(13-15):2306–2318, 2018.

[29] Shuichi Hoshika, Nicole A Leal, Myong-Jung Kim, Myong-Sang Kim, Nilesh B Karalkar, Hyo-Joong Kim, Alison M Bates, Norman E Watkins, Holly A SantaLucia, Adam J Meyer, et al. Hachimoji dna and rna: A genetic system with eight building blocks. *Science*, 363(6429):884–887, 2019.

Chapter 4

Pretransitional Phenomena approaching Liquid Crystal Ordering

4.1 Introduction

Because of their varied, easily controlled and highly temperature-dependent interactions, DNA and RNA oligomers have become a powerful platform to investigate molecular self-assembly. Nucleic acid oligomers are used to explore new concepts for programmable build of nanostructures, to investigate the variety of processes yielding their ordering into nematic, smectic and columnar liquid crystal phases, and to design hydrogels with complex but controlled viscoelasticity. Among these phenomena we are here interested in the processes leading to the onset of the nematic (N) ordering, the simplest among the liquid crystal (LC) phases observed in aqueous solutions of DNA duplexes. Indeed, while DNA nematics have been extensively observed and characterized, studies of pretransitional behavior are still missing. We report here a combined experiments and computer simulations study of the aggregation and local ordering of DNA duplexes both far and in proximity of the isotropic-nematic (I-N) phase boundary. Our results clarify the sharp first-order nature of this phase transition and elucidates the effects of linear ordering on the development of local orientational order.

Pretransitional phenomena have been observed and extensively studied in the context of the I-N phase transition of thermotropic LC, in which they are easily detectable on both sides of the phase boundary. Being a weak first-order transition, N correlations develop in the I phase upon cooling toward the transition, with a divergent behavior having asymptote at a temperature (T) which cannot be reached since the N phase develops. The divergent behavior can however be clearly detected and the divergence T (T^*) determined (typically $T^* \approx T_{IN} + 1°C$, with T_{IN} the I-N transition T). Symmetrically, the N order parameter S decreases upon heating the N phase with a T dependence indicating the existence of a singular point that cannot be reached since it is above, but not far, from the T_{IN}. At T_{IN}, $S \approx 0.4$.

The nematic ordering of DNA duplexes develops only in the presence of end-to-end attractive interactions between duplexes, which, in the case of blunt-ended double helices here considered, is provided by the stacking interactions between terminal paired bases. Because of this combination of orientational order and linear aggregation, DNA LC can be considered as part of the chromonic lyotropic LC family, even though the monomeric units here are elongated duplexes instead of flat polyaromatic molecules, as typical for chromonics. Aggregation and orientation cooperate to the N ordering, as evidenced by computer simulations that show that the length of the aggregates discontinuously grows at the I-N phase boundary from units to tens of monomers! [1, 2]. This well fits with the observation that the N order parameter of chromonics at the I-N transition is much larger ($S > 0.7$) and weakly dependent on T[3] . These observations, which would lead to expecting weak non-critical pretransitional phenomena at the I-N boundary for this class of

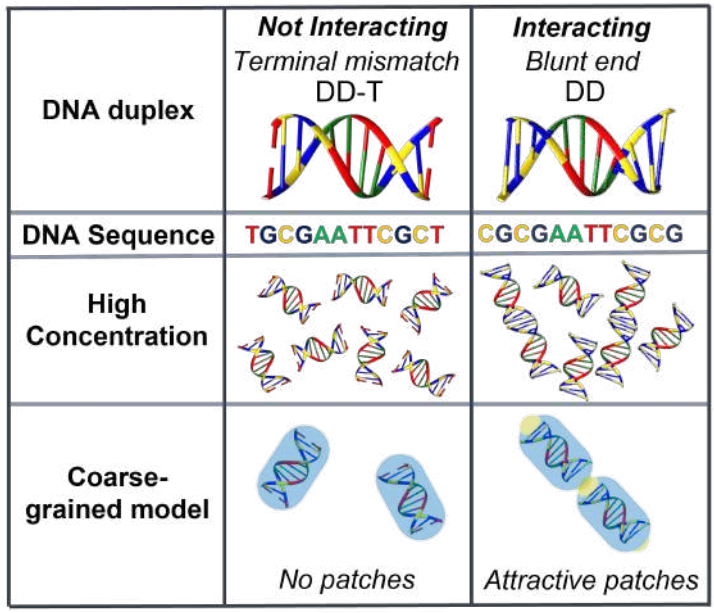

Figure 4.1: Dna duplexes used in SLS experiments. DD-T duplexes have T terminals that prevent interaction. Blunt-ended DD duplexes can interact through stacking interaction. In Monte Carlo computer simulation DNA duplexes are described as cylinder with no patches for DD-T and with attractive patches (yellow) for DD duplexes.

LC, are in conflict with the finding of critical-type N fluctuations in the I phase of chromonic LC, which appear to diverge at a $T^* \approx T_{IN} - 2°C$ [4].

We exploit here the properties of DNA duplexes to access directly the T dependence of both the aggregation number and the orientational correlations via a combination of depolarized static light scattering (SLS) experiments and computer simulations. Indeed, with respect to conventional chromonics, DNA LC whose 'monomers' are the paired duplexes, offer two simplifying properties: (i) the anisotropy of their polarizability $\Delta\alpha = \alpha_{\parallel} - \alpha_{\perp}$, where $\alpha_{\parallel}$ and $\alpha_{\perp}$ are the polarizabilities along and perpendicular to the duplex axis respectively, is large enough to enable signal detection even from solutions of independent (not-aggregated) monomers; (ii) the two terminals of each duplex can be safely considered independent, ensuring that the aggregation process is 'isodesmic', a condition that is not generally obeyed by chromonic LC[5], and certainly not by surfactant-based lyotropic LC.

4.2 Experimental results

Fig.4.2 and Fig.4.3 show the depolarized scattered intensities, I_{HV}, collected at $\theta = 90°$ and $\theta = 135°$ respectively, for both interacting and not interacting DNA duplexes at three different concentrations $c_{DNA} = [132, 237, 313]$ mg/ml, as marked in the legend, and equal salt concentration $c_{NaCl} = 2$M. In panel Fig.4.2(a), the red and blue I_{HV} data clearly show a strictly increasing behaviour until they reach the temperature, $T_{I\to N}$, (vertical dashed lines) at which the isotropic nematic transition occurs. In particular, their I_{HV} at $T=T_{I\to N}$ are ≈ 7 times the corresponding I_{HV} at $T = 70°$C. Such a increase is the signature of aggregation process. Indeed, the bigger the molecule the higher the scattered light. As expected, the $T_{I\to N}$ at $c_{DNA} = 237$ g/l, occurs at a lower temperature compared to $c_{DNA} = 313$ g/l, whereas at $c_{DNA} = 132$ g/l no transition is observed and accordingly its scattered intensity barely increases in even a wider range of temperature.
In order to free the data from T dependencies unrelated to the aggregation process and the development of orientational correlations, I performed similar experiments with not interacting DNA duplexes, $DD-T$, at same DNA and salt concentration as shown in Fig.4.2(b). Their intensities barely increase of a factor 2 in the whole range of temperature explored, and no $I \to N$ transition is observed, because of terminal mismatches that prevent aggregation [6]. Such a behaviour in temperature captures all the features not related to both aggregation process and orientational correlation development. Indeed, these features affect also DD-T behaviour that has a decreasing signal at the highest temperature here explored, whereas the I_{HV} was expected to be constant. At such high temperature, interacting molecules, within the investigate range of c_{DNA}, can be regarded a system of not interacting particles. Such behaviour could be partially due to duplexes melting (causing a change in anisotropy), whose fraction of unpaired bases at T=70°C for all c_{DNA} is less than the 10%, as calculated by standard tools[7].

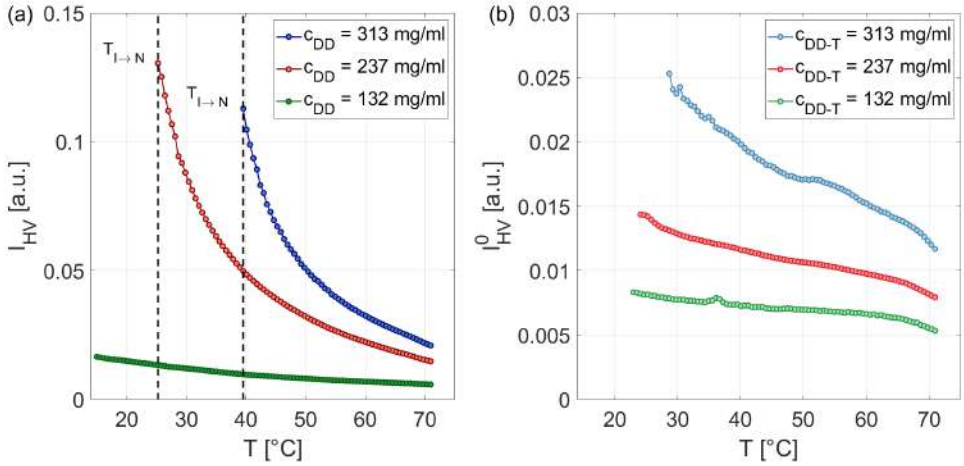

Figure 4.2: Depolarized light scattering intensities for interacting (panel (a)) and not interacting DNA duplexes (panel (b)). Vertical dashed lines correspond to the temperature at which isotropic- nematic transition occurs. DNA concentration are reported in the legend. Data are collected at $\theta = 90°$.

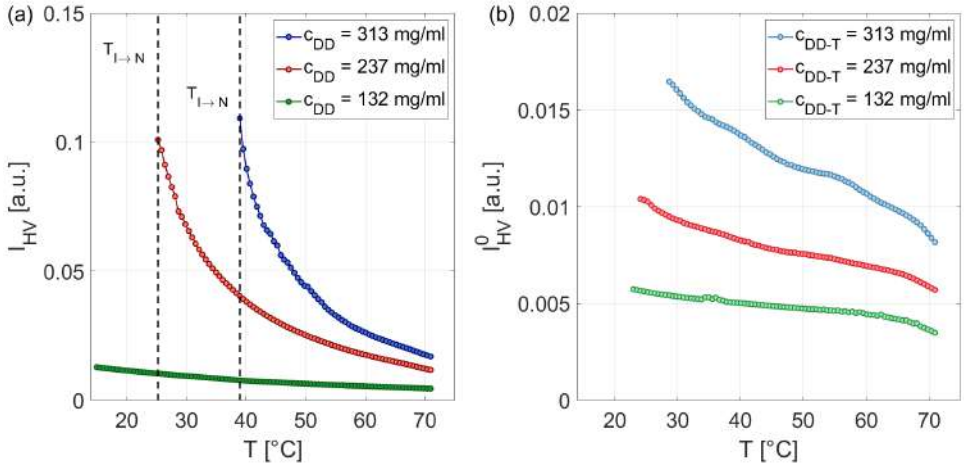

Figure 4.3: Depolarized light scattering intensities for interacting (panel (a)) and not interacting DNA duplexes (panel (b)). Vertical dashed lines correspond to the temperature at which isotropic- nematic transition occurs. DNA concentration are reported in the legend. Data are collected at $\theta = 135°$.

The comparison between Fig.4.2 Fig.4.3 clearly show that the observations described so far for I_{HV} are independent on the scattering angle at which intensities are collected.

If the molecular orientation were random for both linear aggregates and independent particles, the ratio between intensities of interacting (I_{HV}) and

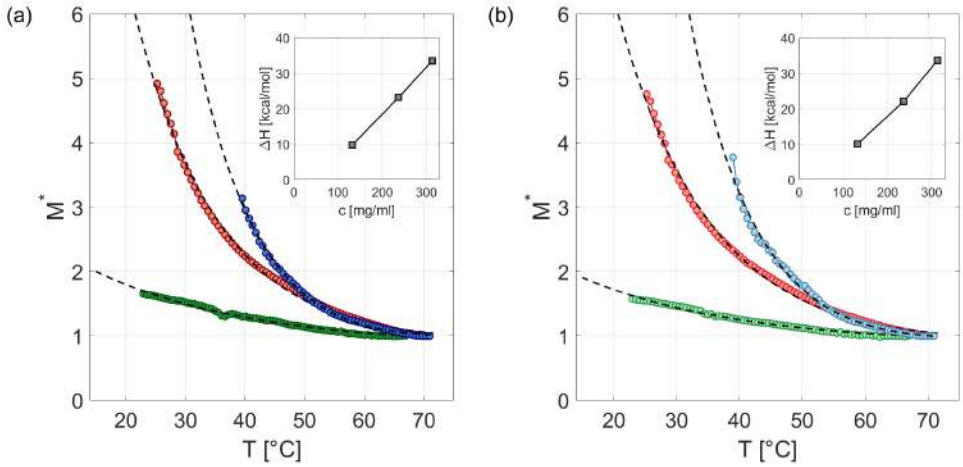

Figure 4.4: Aggregation number M^* calculated as the ratio between the intensities of interacting and not interacting DNA duplexes at corresponding concentration as shown in Fig.4.2 and Fig.4.3. Dashed lines: fit of experimental data according to eq.4.12. Insets: ΔH as a function of DNA concentration and calculated from the fit according to eq.4.12.

noninteracting (I^0_{HV}) duplexes should provide the aggregation number $M^*= I_{HV}/I^0_{HV}$ which is plotted in Fig.4.4. In panel(a) and (b) M^* data are reported for $\theta = 90°$ and $\theta = 135°$ from depolirized intensities shown in Fig.4.2 and Fig.4.3, respectively. Experimental data have been fitted with eq.1.4. For each curve, from the fit it was possible to calculate the interaction energy between duplexes, ΔH, which is shown in the insets in both figures. Although the data are well fitted by the theoretical model, the linear dependence of ΔH on concentration is likely to be a nonphysical result. Indeed, such a result would assert that in the range concentration explored here, the ΔH, i.e. the stacking interaction between duplexes, should change from 10 kcal/mol to 36 kcal/mol, which is a really high and unlikely value for a stacking energy[8]. The guess is that in the aggregation process orientational correlation can develop, so that the increase in scattered intensities is due to DNA duplexes both in linear aggregates and orientationally correlated to them. This contribution can be decoupled with the help of computer simulation. Next section clarify what is the physical quantity that must be calculated in computer simulations.

4.3 Aggregation model and simulation

4.3.1 Depolarized Light Scattering

The scattering geometry is sketched in Fig. 4.5(a), where the scattering plane is xy. The incident light beam propagates along the x direction and it is polarized horizontally (H) respect to the scattering plane, whereas the scattered light intensity is polarized vertically (V) respect to xy plane. The scattered intensity in such configuration is called depolarized scattered intensity I_{HV}.

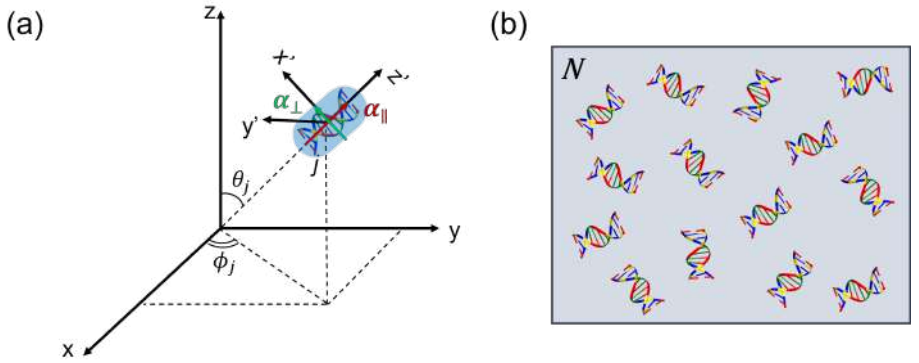

Figure 4.5: (a) Scattering geometry: xy is the scattering plane. The incident light beam propagates along the x direction with horizontal polarization whereas scattered intensity is vertically polarized respect xy. DNA duplex jth is sketched as cylinder with polarizability $\alpha_\parallel$ along its optical axis and $\alpha_\perp$ orthogonal to it. θ_j and ϕ_j describe orientation od cylinder in the xyz reference of frame. Reference of frame $x'y'z'$ is fixed on the cylinder. z' coincides with optical axis, x' lies in the plane formed by zz' and y' is orthogonal to $x'z'$ plane. (b) Aqueous solution of N not interacting DNA duplexes (DD-T).

DNA duplexes are optically anisotropic molecules, whose polarizability tensor has off-diagonal elements. Consequently, the components of the induced dipole moment (eq.2.3) will not usually parallel to the incident field. However, it is possible to choose a set of basis vectors that allow to write the polarizability tensors in a diagonal form, so that along these axes the induced dipole moment and incident field have same direction. An example of basis vectors are the axes of molecule, defined as principal axes of polarizability. DNA molecule can be described as a cylinder which axes of polarizabilities are the axes parallel to its symmetry axis,$\alpha_\parallel$, and the axes perpendicular to $\alpha_\parallel$ in any direction, $\alpha_\perp$, as shown in Fig.4.5(a).
We are interested in evaluating the depolarized scattered intensity I_{HV} in the scattering geometry shown in Fig.4.5(a). Consequently, $n_i = \hat{y}$ and $n_f = \hat{z}$

and eq.2.8 becomes:

$$I_{HV}(\boldsymbol{q}) = \propto \langle \sum_{j,l \in V} \alpha_{yz}^{j}(t)\alpha_{yz}^{l}(t)e^{-i\boldsymbol{q}\cdot(\boldsymbol{r}_l(t)-\boldsymbol{r}_j(t)}\rangle. \tag{4.1}$$

Before calculating α_{yz}, the reference of frame fixed on the cylinder must be introduced in order to work with the polarizability tensor in a diagonal form. Let's choose x' as the axis that lies in the plane formed by the optical axis z' and z, and y' orthogonal to this plane (see Fig.4.5(a)). The angle between x' and and z is $\pi/2$-θ_j. Let's now project the versor $\hat{z}$ and $\hat{y}$ on this reference of frame x', y', z':

$$\hat{z} = \begin{pmatrix} \sin\theta_j \\ 0 \\ \cos\theta_j \end{pmatrix} \tag{4.2}$$

$$\hat{y} = \begin{pmatrix} -\cos\theta_j \sin\phi_j \\ -\cos\phi_j \\ \sin\theta_j \sin\phi_j \end{pmatrix} \tag{4.3}$$

So, it's possible to calculate α_{yz}:

$$\begin{aligned} \alpha_{yz} &= (-\cos\theta_j \sin\phi_j, -\cos\phi_j, \sin\theta_j \sin\phi_j) \begin{pmatrix} \alpha_\perp & 0 & 0 \\ 0 & \alpha_\perp & 0 \\ 0 & 0 & \alpha_\parallel \end{pmatrix} \begin{pmatrix} \sin\theta_j \\ 0 \\ \cos\theta_j \end{pmatrix} \\ &= (\alpha_\parallel - \alpha_\perp)\sin\theta_j \cos\theta_j \sin\phi_j. \end{aligned} \tag{4.4}$$

Independent particles

DNA solution at high temperature are in an isotropic phase. Since short DNA duplexes are *nm*-sized with low axial ratio, it is safe to assume the solution as diluted and that the duplex-duplex orientational correlation length ξ is $\xi << \lambda$. Therefore, the summation in Eq.4.1 can be divided into two terms: the first term for particles whose distance $\Delta r < \xi$ and thus $e^{i\mathbf{q}\cdot[\mathbf{r}_j - \mathbf{r}_l]} \approx 1$; the second term for $\Delta r >> 1$, so that $\langle \alpha_{yz}^{j}\alpha_{yz}^{l} e^{-i\mathbf{q}\cdot[\mathbf{r}_j-\mathbf{r}_l]}\rangle = \langle \alpha_{yz}^{j}\rangle\langle \alpha_{yz}^{l}\rangle\langle e^{-i\mathbf{q}\cdot[\mathbf{r}_j-\mathbf{r}_l]}\rangle$, i.e. the orientation and translation of duplexes far apart are uncorrelated. However, the average α_{yz} over the solid angle is zero, i.e. of $\langle \alpha_{yz}^{j}\rangle = 0$. As a consequence, I_{HV} can be written as:

$$I_{HV} \propto \sum_{j,l \in V} \langle \alpha_{yz}^{j}\alpha_{yz}^{l}\rangle = \sum_{j \in V} \sum_{l=1}^{\bar{N}_j} \langle \alpha_{yz}^{j}\alpha_{yz}^{l}\rangle, \tag{4.5}$$

where $\bar{N}_j$ is the number of scatterers within a distance ξ from particle j that belong to the illuminated volume, too. Notice that in these conditions

the **q** dependence is lost. In the diluted conditions the particles are not interacting each other, and consequently all the cross products vanish and the depolarized scattered intensity for independent duplexes, I^0_{HV} is:

$$I^0_{HV} \propto \sum_{j=1}^{N} \langle \alpha^j_{yz} \alpha^j_{yz} \rangle = \frac{N}{15} (\alpha_{\parallel} - \alpha_{\perp})^2, \tag{4.6}$$

where N counts the duplexes in the illuminated volume V. In the second equality it has been exploited the identity of the particles and the isotropy of the system, by which the orientational distribution of the scatterers is uniform over the solid angle and $\langle \alpha^j_{yz} \alpha^j_{yz} \rangle = \frac{1}{4\pi} \int_0^{\pi} \int_0^{2\pi} (sin^2\theta cos^2\theta sin^2\phi) sin\theta d\theta d\phi = \frac{1}{15}$.

Particles with orientational correlations

Let's now consider the DNA solution approaching the temperature at which the isotropic-nematic transition occurs, T_{IN}. Although, the linear aggregation has a well-developed description enabling quantitative estimates, it is not clear to what extent it is also accompanied by growing orientational correlations. If orientational correlations are not negligible, the cross terms in eq.4.5 cannot be neglected. So, the I_{HV} is:

$$\begin{aligned} I_{HV} \propto & \frac{N}{15}\beta^2 + \beta^2 \sum_{j=1}^{N} \sum_{l=1, l\neq j}^{\bar{N}_j} \langle sin\theta_j cos\theta_j sin\phi_j sin\theta_l cos\theta_l sin\phi_l \rangle = \\ = & \frac{N}{15}\beta^2 + N\beta^2 \sum_{l=1, l\neq j}^{\bar{N}_j} \langle sin\theta_j cos\theta_j sin\phi_j sin\theta_l cos\theta_l sin\phi_l \rangle, \end{aligned} \tag{4.7}$$

where $\beta = \alpha_{\parallel} - \alpha_{\perp}$ is the anisotropy of particle polarizability. This equation can be further simplified by expressing the orientation of the particle l by the polar coordinates θ'_l, ϕ'_l with respect to the reference frame (x',y',z') fixed to the particle j, as sketched in Fig.4.6. The unit vector along the l particle is:

$$\boldsymbol{u}_l = sin\theta'_l cos\phi'_l \boldsymbol{u}_{x'} + sin\theta'_l sin\phi'_l \boldsymbol{u}_{y'} + cos\theta'_l \boldsymbol{u}_{z'} \tag{4.8}$$

where the unit vectors of the (x',y',z') system can be expressed in terms of the lab frame (x,y,z)[]:

$$\begin{aligned} \boldsymbol{u}_{x'} &= (\cos\theta_j \cos\phi_j, \cos\theta_j \sin\phi_j, -\sin\theta_j) \\ \boldsymbol{u}_{y'} &= (-\sin\phi_j, \cos\phi_j, 0) \\ \boldsymbol{u}_{z'} &= (\cos\theta_j \sin\phi_j, \sin\theta_j \sin\phi_j, \cos\theta_j), \end{aligned} \tag{4.9}$$

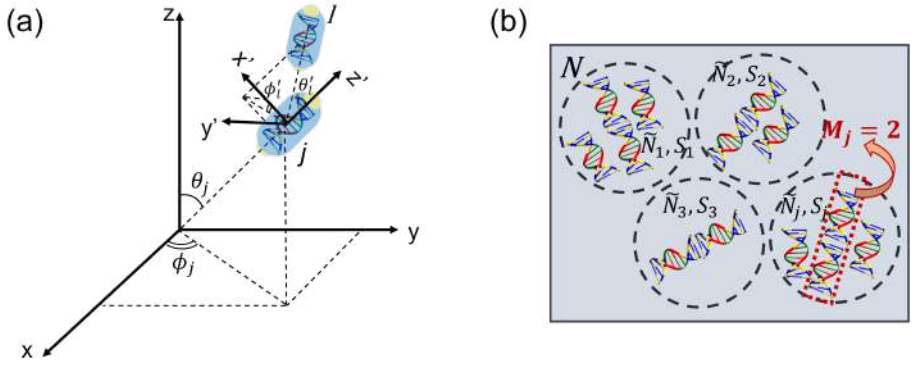

Figure 4.6: The scattering geometry, orientation of j particle and the reference of frame $x'y'z'$ is identical to the ones described in Fig. 4.5. The orientation of particle l is described by θ'_l and ϕ'_l in the $x'y'z'$ reference of frame. (b) Aqueous solution of N interacting DNA duplexes (DD) grouped in domains of N_j molecules orientationally correlated described by local order parameter S_j. M_j is the aggregation number of duplexes in the j domain.

can be expressed in terms of the lab frame (x,y,z) to extract the two quantities $cos\theta_l = \boldsymbol{u}_l \cdot (0,0,1)$ and $sin\theta_l sin\phi_l = \boldsymbol{u}_l \cdot (0,1,0)$. Eq.4.7 can be thus be written as:

$$
\begin{aligned}
I_{HV} &= \frac{N}{15}\beta^2 + \\
&+ \frac{N}{4\pi}\beta^2 \sum_{l=1,l\neq j}^{\bar{N}_j} \iint_{\Omega} \sin\theta_j \, d\theta_j \, d\phi_j \langle \sin\theta_j \cos\theta_j \sin\phi_j (\boldsymbol{u}_l \cdot (0,0,1))(\boldsymbol{u}_l \cdot (0,1,0)) \rangle = \\
&= \frac{N}{15}\beta^2 \Big(1 + \sum_{l=1,l\neq j}^{\bar{N}_j} \langle \frac{3\cos^2\theta'_l - 1}{2} \rangle \Big),
\end{aligned}
\tag{4.10}
$$

where, in the last equation, we assumed that the distribution in ϕ'_l is flat. The last equality makes clear the physical meaning of the result. The depolarized scattered intensity of DNA duplexes orientaltionally correlated is given by the sum of two contributions: the first term is the I_{HV} for independent particles; the second term is the sum of small nematic clusters, described by their local nematic order parameter, composed by all particles l orientationally correlated to the j particle (see Fig.4.6b).

4.3.2 Monte Carlo simulation

The comparison between experiments and the model of aggregation led to nonphysical results. While the linear aggregation has a well-developed description enabling quantitative estimates, it is not clear to what extent it is also accompanied by growing orientational correlations. Such information can be quantified from computer simulations. To this aim, the first step consists on understanding what is the quantity to compute. In Fig.4.4 the experimental I_{HV} for the interacting duplexes have been divided by the experimental I_{HV} for independent scatterers at $c_{DD-T} = 132$ g/l. So, if orientational correlations occur during aggregation process, the quantity sought is exactly the ratio between eq.4.6 and eq.4.10:

$$M^* = \frac{I_{HV}}{I^0_{HV}} = 1 + \sum_{l=1, l\neq j}^{\bar{N}_j} \langle \frac{3\cos^2\theta'_l - 1}{2} \rangle. \tag{4.11}$$

In computer simulation, the interacting DNA duplexes are described as hard cylinders (HCs) with aspect ratio $X_0 = L/D$=2, and with attractive patches on both bases of the cylinder, as sketched in Fig.4.6(a). The attractive sites on two distinct cylinders interact each other with a square well potential within a range $\delta = 0.25D$ and binding energy u_0, which is independent on aggregation size. Metropolis Monte Carlo simulations have been performed on interacting HCs at the same three different experimental concentration c = [132, 237, 313] g/l as a function of adimensional temperature $T^*= k_BT/u_0$. M^* has been calculated according to eq.4.11 with Legendre polynomials of order 2, $P_2(\theta)$, in a neighbourhood of half simulation box around each particle j, as sketched in Fig.4.6(b). M has been calculated from the computed total energy of the system E, as $M = 1/(1 + E/(nu_0))$, where n is the number of cylinders in the aggregate. The total number of HCs involved in simulation is $N \approx 10^4$. Further details of computer simulation are reported in [9]. The outcomes of simulation are reported in Fig.4.7. In particular, Fig.4.7(a) and (b) show the behaviour of the real aggregation number M and M^*, respectively as a function of adimensional temperature T^* for all the experimental concentration explored. Such results are unified in panel(c) to highlight the not trivial relation between M and M^*. Indeed, for the two highest concentrations, in proximity of $I \rightarrow N$ transition, the contribution of orientational correlations is so strong that $M^* \approx 2M$. Dashed lines are polynomial fit that allows to get the sought relation between M and M^*, so that the experimental M^* can be converted in the experimental aggregation number. If the only the particles orientationally correlated with particle j are those participating in an aggregate and if the aggregates are nearly straight, $M^* = M$, and the normalized depolarized intensity is a measure of the degree of aggregation.

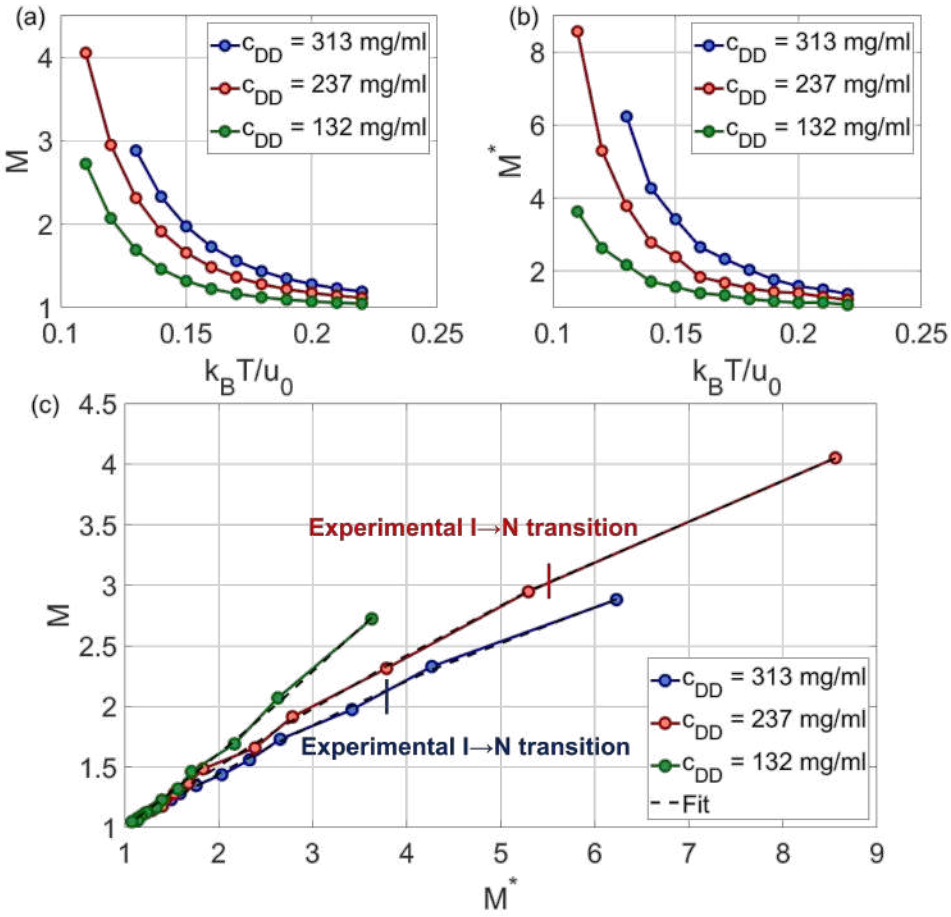

Figure 4.7: Outcomes of Metropolis Monte Carlo computer simulations of hard cylinders with interacting patches, at three different concentration c as reported in the legends. Panel(a) and (b) show the behaviour of the aggregation number M and M^*, respectively, as a function of adimensional temperature $T^* = k_bT/u_0$. Panel(c) show the behaviour of M as a function of M^* calculated at the same T^*. Dashed lines are polynomial fits.

4.4 Discussion

The simulation results shown in Fig.4.7 allow us to calculate the aggregation number M from experimental M^* shown in Fig. 4.4. The results are shown in Fig.4.8, where M data are fitted with eq.1.4 providing the energies ΔH plotted in the insets, that are lower than the corresponding ΔH calculated in fig.4.4. In particular, ΔH for c =132 g/l slightly decreases accordingly to the outcome of MC simulation where the effect of orientational correlation on M for such low c is marginal. Differently for the higher concentrations ΔH

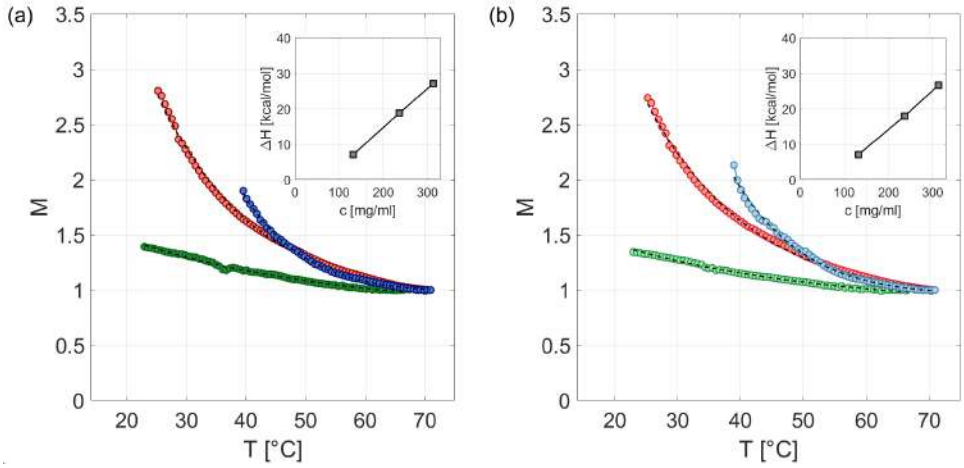

Figure 4.8: Aggregation number M calculated from experimental M^* reported in Fig.4.4 and converted in M through computer simulation results shown in Fig.4.7. Data shown in panel (a) and panel (b) correspond to data collected at θ=90° and θ=135°,respectively. Coloured lines correspond to DD concentration values as reported in Fig,4.4. Insets: ΔH calculated from the the fit (dashed lines) according to 1.4.

values are decreased by $\approx 5kcal/mol$. In Fig.4.9 the M data are fitted with eq.1.4 but with ΔH as global free parameter providing ΔH = 18 kcal/mol for both angles. Although the values of ΔH for DNA solution that show I-N transition are still too high, their values are decreased thanks to simulation results and ΔH = 18 kcal/mol correspond to almost the double of expected value. Moreover, the contribution of orientational correlation computed by MC simulation is likely to be underestimated. Indeed, MC simulation didn't reveal any transition into nematic phase, whereas in experiments such transition occurs at lower M^* (vertical lines in Fig.4.7). So, a further comprehension on such disagreement could lead to lower ΔH comparable with the ΔH of DNA solution at c = 132 g/l (inset Fig.4.8).

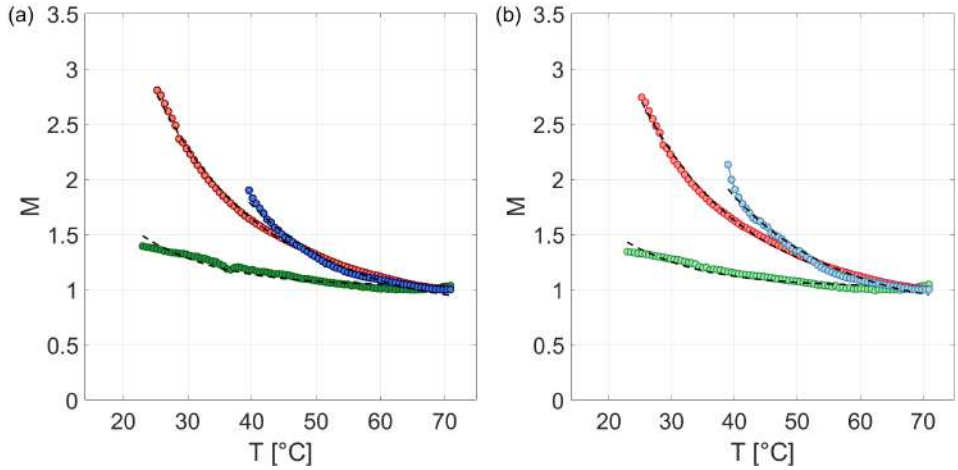

Figure 4.9: Aggregation number M calculated from experimental M^* reported in Fig.4.4 and converted in M through computer simulation results shown in Fig.4.7. Data shown in panel (a) and panel (b) correspond to data collected at θ=90° and θ=135°,respectively. Coloured lines correspond to DD concentration values as reported in Fig,4.4. Dashed lines: global fit through 1.4, where the only global free parameter is the enthalpy ΔH. For both θ=90° and θ=135°, ΔH =18 kcal/mol.

Future directions

In order to understand better to what extent orientational correlation develop, an interesting experiment involves DNA duplexes that can form only dimer. Indeed, even at high concentration such a system don't show any LC transition and two plateau values are excepted at high and low T. The plateau at high T regime corresponds to monomers condition, whereas the plateau at low T, where mostly dimer are formed, should give us the information about M^* which inevitably corresponds to aggregation number M=2. The comparison with MonteCarlo simulation results and the behaviour of experimental M^* as a function of T of the dimer solution at different concentration should help in better understanding the activation energy results and dependence on c. The latter result could make possible to extrapolate a dependence of ΔH on T, that so far it was assumed constant over all the aggregation process.

4.5 Material and Methods

4.5.1 Data Analysis

Fig.4.10 shows the steps in the analysis of experimental depolarized scattered intensities to get M^* and ΔH as discussed in section4.2. Experimental I_{HV} shown in panel(a) are firstly smoothed, averaging data on a sliding window of

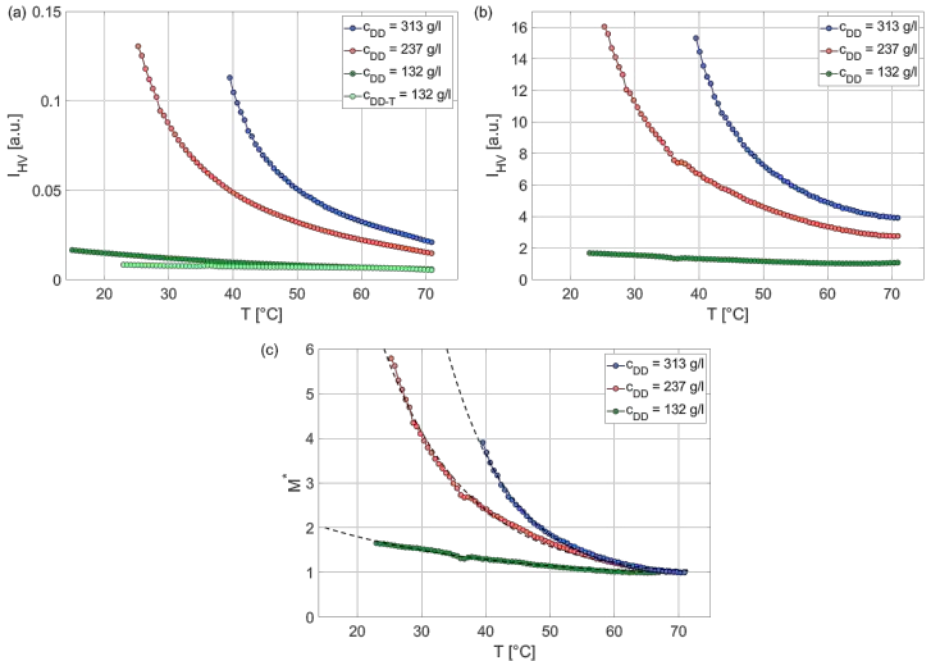

Figure 4.10: Steps in analysis of depolarized scattering intensities, I_{HV}, of data collected at $\theta = 90°$. (a) Depolarized scattering intensities after smoothing of DD and DD-T duplexes as reported in the legend. (b) I_{HV} of interacting molecules divided by the I_{HV} of not interacting molecules at $c_{DD-T} = 132$ g/l shown in panel (a). (c) M^* of interacting DNA duplexes calculated as normalization of I_{HV} in panel (b) by their corresponding value at high T. Dashed lines: fit according to 4.12.

width 0.5°C. Secondly, in panel (b), I_{HV} of the interacting duplexes are normalized by the I_{HV} of not interacting duplexes to get rid of contributions not related to both aggregation and orientational correlation. Such normalization allows to get a plateau at high T, as initially expected in raw data. Indeed at high T no aggregation process occurs and I_{HV} is expected to constant if other phenomena are not involved. Finally, in panel (c) the I_{HV} plotted in panel(b) are normalized by the corresponding plateau value at high T which correspond to the diluted condition where no aggregates are formed, and so $M^* = M$ =1. M^* data are fitted by:

$$y(T) = a_1(1 + \sqrt{1 + a_2(e^{a_3/T})}), \tag{4.12}$$

where T are in Kelvin, a_i with $i = 1, 2, 3$ are free parameters. In particular, ΔH(kcal/mol) can be calculated from a_3 parameter by $\Delta H = a_3 R/4186$, where R is the gas constant. The fits of aggregation number M shown in Fig.4.9, are performed with eq.4.12 where the a_3 parameter is differently set

as a global parameter for all the experiments, whereas a_1 and a_2 are free parameters for each curve.

4.5.2 DNA sample preparation

DD and DD-T were firstly diluted and to a concentration equals to 50 mM nucleotides and dialysed in a a buffer at $c_{NaCl} = 25mM$. After lyophilising, DNA was characterized by UV- absorbance measurement with UV-VIS 300 ThermoScientific spectrophotometer, that revealed $m_{DNA}^{DD} \approx 70\% m_{tot}^{DD}$ and $m_{DNA}^{DD-T} \approx 87\% m_{tot}^{DD-T}$, where m_{tot} is weighted mass. Assuming that the remaining part is entirely salt, four different Hilgenberg round capillaries of 2.4 mm inner diameter and length 30 mm were prepared. In particular, t to $c_{DD} = 313 - 237 - 132 \quad mg/ml$ and $C_{Nanucleotides} a 2M$ were filled with 9.3-7.7-4.9 mg, respectively, of lyophilised DD. Successively water was added in order to achieve the aforementioned c_{DD} and $c_{NaCl} = 2M$ with a total volume in the range $21 - 26\mu l$. The fourth capillary was filled similarly with 11.9 mg of lyophilised DD-T. It was initially added water and salt to get the final concentration $c_{DD-T} = 313 g/l$ and $cNaCl = 2M$. It was successively diluted to $c_D D - T = 237,132 mg/ml$ keeping constant $c_{NaCl} = 2M$.

Bibliography

[1] Cristiano De Michele, Tommaso Bellini, and Francesco Sciortino. Self-assembly of bifunctional patchy particles with anisotropic shape into polymers chains: Theory, simulations, and experiments. *Macromolecules*, 45(2):1090–1106, 2012.

[2] Tatiana Kuriabova, MD Betterton, and Matthew A Glaser. Linear aggregation and liquid-crystalline order: comparison of monte carlo simulation and analytic theory. *Journal of Materials Chemistry*, 20(46):10366–10383, 2010.

[3] John Lydon. Chromonic liquid crystals. *Handbook of liquid crystals*, pages 1–45, 2014.

[4] Yu A Nastishin, H Liu, SV Shiyanovskii, OD Lavrentovich, AF Kostko, and MA Anisimov. Pretransitional fluctuations in the isotropic phase of a lyotropic chromonic liquid crystal. *Physical Review E*, 70(5):051706, 2004.

[5] Peter J Collings, Joshua N Goldstein, Elizabeth J Hamilton, Benjamin R Mercado, Kenneth J Nieser, and Margaret H Regan. The nature of the assembly process in chromonic liquid crystals. *Liquid Crystals Reviews*, 3(1):1–27, 2015.

[6] Simone Di Leo, Marco Todisco, Tommaso Bellini, and Tommaso P Fraccia. Phase separations, liquid crystal ordering and molecular partitioning in mixtures of peg and dna oligomers. *Liquid Crystals*, 45(13-15):2306–2318, 2018.

[7] Nupack. http://www.nupack.org/partition/new.

[8] Cristiano De Michele, Lorenzo Rovigatti, Tommaso Bellini, and Francesco Sciortino. Self-assembly of short dna duplexes: from a coarse-grained model to experiments through a theoretical link. *Soft Matter*, 8(32):8388–8398, 2012.

[9] Khanh Thuy Nguyen, Francesco Sciortino, and Cristiano De Michele. Self-assembly-driven nematization. *Langmuir*, 30(16):4814–4819, 2014.

Chapter 5

Molecular diffusion in structured fluids

5.1 Introduction

In this chapter it is investigated the molecular diffusion in structured fluids, where the participation of individual molecules to the structure is typically transient, a condition which enables them to diffuse. A general description of such diffusive process is however still elusive mainly because of the complex interplay of attractive and repulsive interactions among the molecular species involved in the structure. Lyotropic chromonic liquid crystals (CLC) formed by DNA oligomers are an appealing model for crowded, self-assembled biomolecular fluids, where diffusion can be investigated. In CLC the order forms as a result of hierarchical molecular self-assembly[1, 2], whose features are controlled by DNA sequence design. Specifically, experiments have been performed on isotropic and nematic CLC phases formed by DNA duplexes 12-base-long oligomers having different hybridization pattern (see Fig.5.1), a choice that enabled us to compare the behaviour of distinct systems that share the same shape and electric charge of the monomers. Using this design, we could thus compare diffusion in concentrated solutions of purely repulsive duplexes with terminals that prevent any attractive interactions, to the diffusion in solutions of duplexes with controlled end-to-end attraction. DNA duplexes used in this work are shown in Fig.5.1. The repulsive duplexes ,named as FT,have TT terminals that forbid attractive interaction. The attractive duplexes are named DD and sDD, attractively interact by stacking interactions and by self complementary overhangs, respectively. These systems are investigated by a combination of FRAP (Fluorescence Recovery After Photobleaching) and static and dynamic, polarized and depolarized light scattering, from which it was possible to determine the rotational and translational, self and collective diffusion processes. In particular, I carried out the dynamic light scattering (DLS) experiments, whereas FRAP experiments were performed by Dr. Marco Todisco and they are here reported to help in understanding DLS results and vice versa.
In the first section, the theoretical framework is introduced for polarized and depolarized field correlation functions. In the second section, the results of DLS and FRAP experiments are reported for the following DNA duplexes (see Fig.5.1) and concentration c: nonlinear interacting FT duplexes at c = 300 and 200 g/l; blunt ended DD duplexes, that interact by stacking interaction, at c = 300-200-100g/l; sDD duplexes that interact through self-complementary overhangs at c = 300g/l. For each experiment are reported the squared field correlation functions, and the corresponding decay times and diffusion coefficients calculated from the fits. The last sections are dedicated to the discussion of results and to materials, whereas the experimental setup is discussed in section2.1.6.

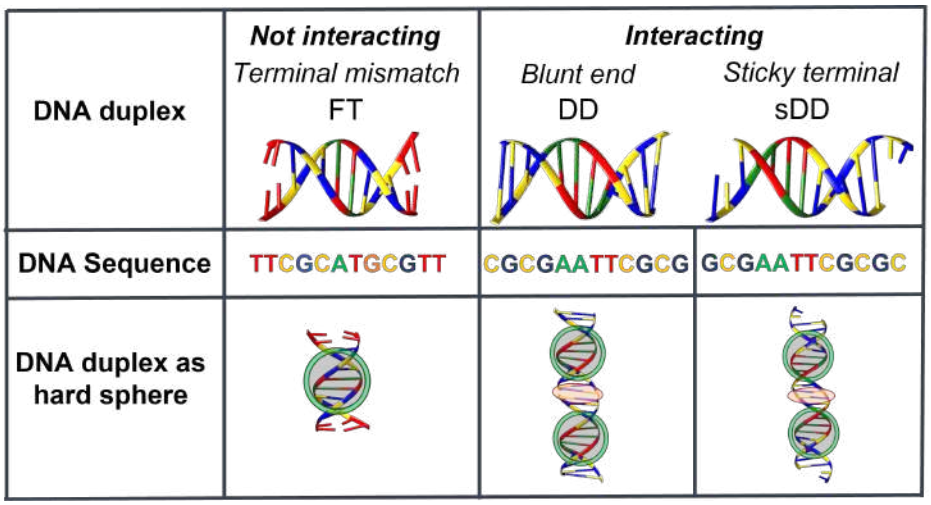

Figure 5.1: Dna duplexes used in DLS experiments. FT duplexes have TT terminals that prevent interaction. Blunt-ended DD duplexes can interact through stacking interaction. sDD duplexes can sticky each other because of self-complementary terminals. DNA duplexes can be described as spheres with radius equivalent to its volume (grey sphere) and Debye length (green sphere). Red circles marks the attractive interaction.

5.2 Polarized and depolarized DLS

Scattered fields by a single anisotropic particle in depolarized (HV), with $\boldsymbol{n}_i$=H and $\boldsymbol{n}_f$=V, and polarized (VV) with $\boldsymbol{n}_i$=$\boldsymbol{n}_f$ =V, configuration can be calculated according to both eq.2.4 and scattering geometry in Fig.4.5(a) as:

$$E_{HV} = \frac{-k_z^2 E_0}{4\pi\epsilon R} e^{i(k_z R-\omega t)} e^{i\boldsymbol{q}\cdot\boldsymbol{r}(t)} \alpha_{yz}(t) \tag{5.1a}$$

$$E_{VV} = \frac{-k_z^2 E_0}{4\pi\epsilon R} e^{i(k_z R-\omega t)} e^{i\boldsymbol{q}\cdot\boldsymbol{r}(t)} \alpha_{zz}(t), \tag{5.1b}$$

where α_{yz} is given by eq.4.4, whereas α_{zz} can be calculated analogously as:

$$\begin{aligned}\alpha_{zz} &= (\sin\theta_j, 0, \cos\theta_j) \begin{pmatrix} \alpha_\perp & 0 & 0 \\ 0 & \alpha_\perp & 0 \\ 0 & 0 & \alpha_\parallel \end{pmatrix} \begin{pmatrix} \sin\theta_j \\ 0 \\ \cos\theta_j \end{pmatrix} \\ &= \alpha_\parallel \cos^2\theta_j + \alpha_\perp \sin^2\theta_j.\end{aligned} \tag{5.2}$$

So, for dilute system and assuming translation and rotation of molecules as statistical independent, according to eq.2.15 the depolarized and polarized field correlation functions are[3, 4]:

$$G_{1,HV}(\tau) = \frac{\beta^2}{15} \langle N \rangle e^{-6D_{rot}\tau - D_T^{self} q^2 \tau} \tag{5.3}$$

$$G_{1,VV}(\tau) = \langle N \rangle \left(\frac{4}{45} \beta^2 e^{-6D_{rot}\tau - D_T^{self} q^2 \tau} + \alpha^2 e^{-D_T^{coll} q^2 \tau} \right), \qquad (5.4)$$

where $\alpha = (\alpha_{\parallel} + 2\alpha_{\perp})/3 - v_p(n^2 - 1)$ is the average difference between particle and solvent polarizabilities, v_p is the volume of the particle. It's important to notice that $G_{1,HV}$ is measurable only if particle polarizabilty is anisotropic, i.e. $\beta \neq 0$.

5.2.1 Short DNA duplexes as hard spheres

Short DNA duplexes can be described as hard spheres, as depicted in Fig.5.1, with a radius R given by the sum of two contributions: the radius of a sphere, R_s= 1.44nm, equivalent to the volume of a cylinder of length L= 4nm and diameter d=2nm[5]; Debye radius $\lambda_D \approx 0.3\sqrt{I}$=0.21nm, where I is the ionic strength equals to 2M of NaCl. So, the equivalent radius sphere is R=1.65nm.

5.3 Experimental results

Eq.5.3 and 5.4 highlight three different diffusion coefficients. They can be independently measured by using three different techniques: (1) FRAP measurements allow to measure the self diffusion coefficient; (2) depolarized light scattering experiments allow to quantify the rotational diffusion coefficient as described by eq.5.3, where D_T^{slef} is determined by FRAP experiments; (3) polarized light scattering experiments allow to measure collective diffusion coefficient as described by eq.5.4, where D_{rot} and D_T^{self} are determined by depolarized light scattering and FRAP experiments, respectively. This procedure has been used to measure the diffusion of solution of interacting duplexes. Differently, for the nonlinear interacting duplexes at any concentration and on the whole rage of temperature explored, the low contrast and fast decay time of depolarized field correlation functions make impossible the analysis described above. Consequently, for FT duplexes the analysis of polarized correlation functions provided both rotational and diffusion coefficients, whereas D_T^{self} is still quantified by FRAP experiments.

5.3.1 Scattering from nonlinear interacting DNA particles

Fig.5.2(a) shows the experimental squared field correlation functions in VV configuration at $\theta = 90°$ for nonlinear interacting FT DNA duplexes at high concentration $c = 300$ g/l. $G^2_{1,VV}$ were obtained from experimental intensity field correlation function as described in section5.5.1. For sake of clarity, in Fig.5.2(a), only three different temperatures $T = 6$, 12.5 and 20°C (from blue

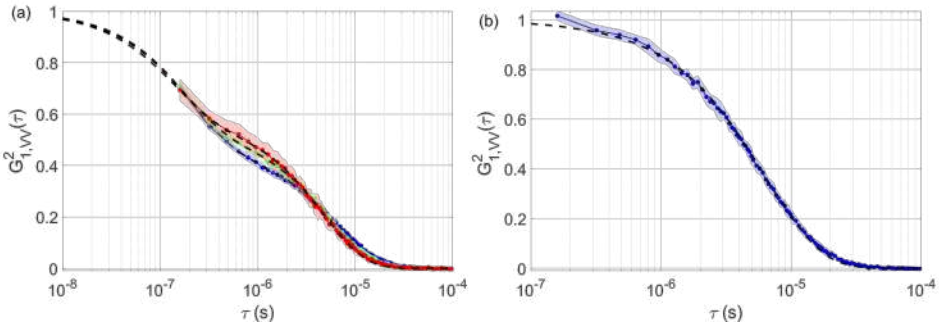

Figure 5.2: (a)Squared field correlation functions $G^2_{1,VV}$ at $\theta = 90°$ in VV configuration at three different $T = 6$ (blue dots), 12.5(green dots) and 20(red dots)°C for FT DNA duplexes at $c = 300$ g/l. Experimental data are fitted with eq.5.13 (dashed lines). (b)Squared field correlation functions $G^2_{1,VV}$ in VV configuration at $T = 6$°C or FT DNA duplexes at $c = 200$ g/l. Experimental data are fitted with a single exponential decay (dashed line). Shaded region marks experimental uncertainty.

to red) are plotted. In all the $G^2_{1,VV}$ data, it is evident a fast exponential decay, whose relative amplitude and decay time decrease at increasing T and they completely disappear at $T = 30$°C (data not shown). Differently, the decay time of the slower exponential decay doesn't show an evident dependence on T. Diluting the DNA solution down to $c = 200$ g/l the fast decay is not anymore observable even at the lowest temperature explored, which is T = 6°C, as shown in Fig.5.2(b). Obliviously, in Fig.5.2(b), the contrast of $G^2_{1,VV}$ is not really equals to 1, because we cannot experimentally investigate lower delay times τ, where orientational decay occurs. Fig.5.3 shows both the decay times calculated from the fit of the squared field correlations with eq.5.13 and the decay times corresponding to self diffusion coefficients, D_T^{self}, measured in FRAP experiments, which are calculated as $\tau_{self} = 1/(D_T^{self} q^2)$. The difference of ≈ 4 order of magnitudes between τ_{self} and τ_{rot} makes clear that first exponential in eq.5.4 decorrelates because of the fast rotational diffusion. Fig.5.4 shows the adimensional D/D_0, corresponding to the decay times reported in Fig.5.3. The experimental translational and rotational diffusion coefficients, D, are divided by corresponding translational and rotational diffusion coefficients of a single particle, D_0, given by eq.2.11 and eq.2.12, respectively, where the radius $R = 1.52$ nm and the viscosity $\eta(T)$ are measured by FRAP experiments. Collective diffusion (dark and light blue circles) is constant over the range of temperature explored for both concentration and their values are reported in Tab.5.1. Regarding FT DNA duplexes as hard spheres (HS) with radius $R = 1.65$ nm, the adimensional diffusion

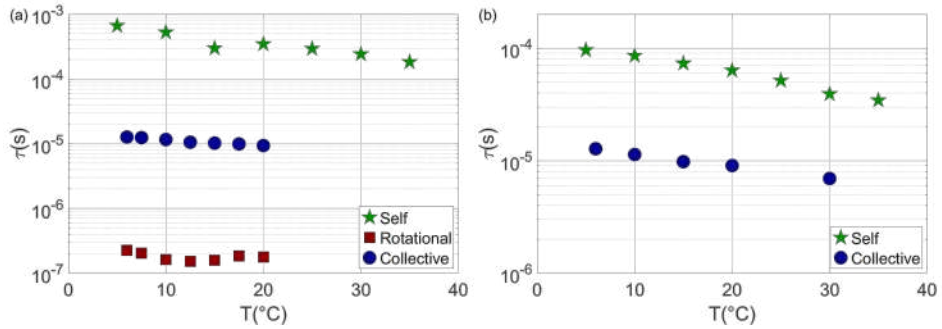

Figure 5.3: Comparison of decay times calculated from DLS and FRAP experiments for FT at $c = 300$ g/l in panel (a) and $c = 200$g/l in panel (b). Symbols: green stars are calculated from diffusion coefficient of FRAP experiments; blue circles and green squares are rotational and collective decay times respectively, calculated from the fit of experimental data with eq.5.13.

coefficient is described by[6]:

$$\frac{D_T^{coll}}{D_0} = 1 + \lambda\phi, \tag{5.5}$$

where ϕ is the volume fraction ($\phi = 0.47$ and 0.31 for $c = 300$ and 200 g/l respectively), and $\lambda = 1.45$. So, for the experimental volume fractions, the predicted D_T^{coll}/D_0 according to eq.5.5 are 1.68 and 1.45, respectively. The former value is almost compatible with the experimental $D_T^{coll}/D_0 = 1.78 \pm 0.08$. Such a small difference is likely to be due to repulsive interactions that increase λ value[6]. Indeed, DNA particles are negatively charged, partially screened by ions, and so their interaction is a combination of steric (as hard spheres) and repulsive contributions. Self diffusion coefficients from FRAP experiments (dark and light green stars) are nicely described by[7]:

$$\frac{D_T^{self}}{D_0} = \frac{(1-\phi)^3}{(1 + 1.5\phi + 2\phi^2 + 3\phi^3)}, \tag{5.6}$$

which takes into account also hydrodynamic interactions. Experimental and theoretical predictions of adimensional diffusion coefficients calculated according to 5.6 are reported in Tab.5.1. Differently, from collective results, the experimental D_T^{self}/D_0 scale correctly with ϕ and are in accordance with theoretical predictions. Rotational diffusion coefficient for hard spheres involving two and three body hydrodynamic interaction is[8]:

$$\frac{D^{rot}}{D_0} = 1 - 0.631\phi - 0.726\phi^2. \tag{5.7}$$

Experimental values differ of one order of magnitude from predicted ones according to eq.5.7. Such a phenomena can be described in terms of effective

ϕ	$D_T^{coll,exp}/D_0$	$D_T^{coll,th}/D_0$	$D_T^{self,exp}/D_0$	$D_T^{self,th}/D_0$
0.47	1.78±0.08	1.68	0.048±0.009	0.06
0.31	1.79±0.08	1.45	0.24±0.03	0.19

Table 5.1: Adimensional collective and rotational diffusion coefficient and corresponding theoretical predictions for nonlinear interacting DNA duplexes at two different volume fraction ϕ.

viscosity, η_{eff}. By fitting all $G^2_{1,VV}$ data with eq.5.14, it's possible to get the effective viscosity that molecules feel rotating as $\eta_{eff} = \gamma\eta_{water} \approx 42\eta_{water}$. The black dashed line in Fig.5.4 is the global fit the gives the result of $1/\gamma$.

5.3.2 Scattering from interacting DNA duplexes

DD molecules are blunt end duplexes that interact each other by stacking interaction. Dynamic diffusion of DD have been investigated at high enough concentration ($c \approx 300g/l$) to show liquid crystal transition into a nematic LC phase above its characteristic T_{IN}. DLS experiments have been performed until T_{IN} is reached, whereas FRAP measurements could have been performed also in LC phase. Phase diagrams of DD molecules are thoroughly discussed in[9].

DD at high concentration

Although the theoretical model introduced in section5.2 holds for nonlinear interacting particles, such a model can be applied as approximation to interacting particles introducing the stretching exponent α. This parameter takes into account of polidispersity of cluster size due to possible aggregation process. The analysis of the intensity and squared field correlation functions is thoroughly discussed in section5.5.1. Fig.5.5 shows experimental squared field correlation functions for DD duplexes at $c = 300$ g/l and at three different temperatures. $G^2_{1,HV}$ are well fitted with a stretched exponential decay with the stretched exponent $\alpha = 0.8$ (see eq.5.12). The stretch exponent reveal a polidispersity in DNA solution, that is likely to be due to aggregation process and consequently cluster formation approaching the temperature T_{IN} at which nematic transition occurs. The fast single decay time, which differs 2-3 order of magnitude from self decay time (see Fig.5.6) and its independence on the scattering angle θ (data not shown here) confirm it to be due to rotational diffusion in accordance with the theoretical model described by eq.5.3. Differently from nonlinear interacting duplexes, it' s evident the strong dependence of rotational decay times on T. Indeed, at high T, τ_{rot} for both FT and DD duplexes are almost equal, and the latter turns slower and

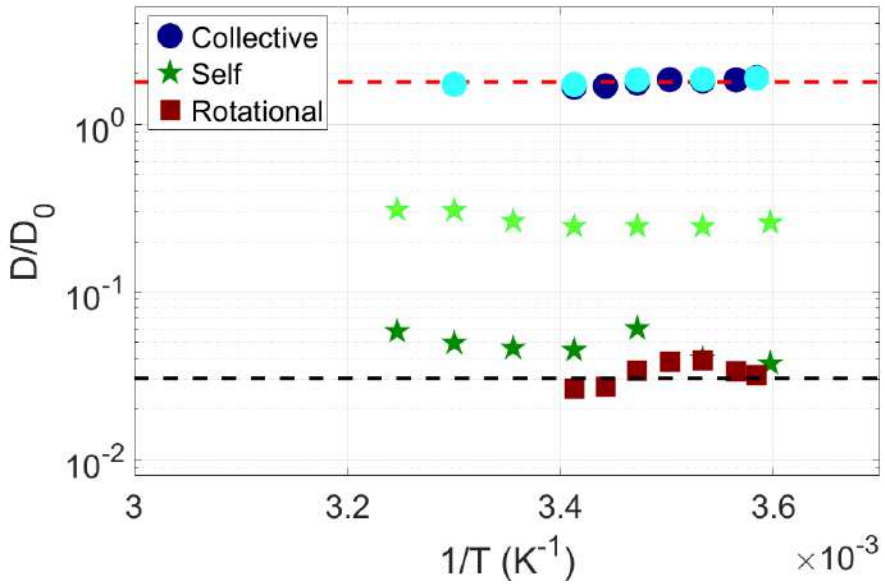

Figure 5.4: Adimensional diffusion coefficients D/D_0 vs. $1/T$. Symbols: stars correspond to self diffusion coefficients calculated in FRAP experiments; squares and circles are rotational and collective diffusion coefficients from DLS experiments calculated from fits with eq.5.13. Dark and light colours correspond to FT at $c = 300$ and 200 g/l, respectively. Red dashed line shows the linear fit of collective diffusion coefficients and it provides $D_{coll}/D_0 \approx 1.78$ for both FT concentrations. Black dashed line is γ calculated as global parameter with eq.5.13 for all experimental $G^2_{1,VV}$ curves for all T.

slower approaching T_{IN}. Such behaviour can be nicely explained in terms of growing cluster of aggregated duplexes and development of orientational correlation that hinder rotational diffusion more and more, in accordance with SLS results discussed in section4.4.

The $G^2_{1,VV}$ in Fig.5.6(b) at T=35° and T=42° show a very similar decay to the corresponding $G^2_{1,HV}$, whereas at T=52° a slower decay time clearly appears. Fig.5.6 nicely displays rotational and collective decay times and how they tend to collapse at decreasing T. Although, self diffusion decay times are 2 order of magnitudes slower than rotational ones, they have a similar behaviour on T and surprisingly the trend in T is unaffected by the insurgence of LC phase. The physical meaning of such dependence on T can be appreciated by looking at the adimensional diffusion coefficients shown in Fig.5.7. Indeed, both self and rotational diffusion exhibit an activated process that can be described by:

$$\frac{D}{D_0} = e^{-\frac{\Delta G}{RT}}, \tag{5.8}$$

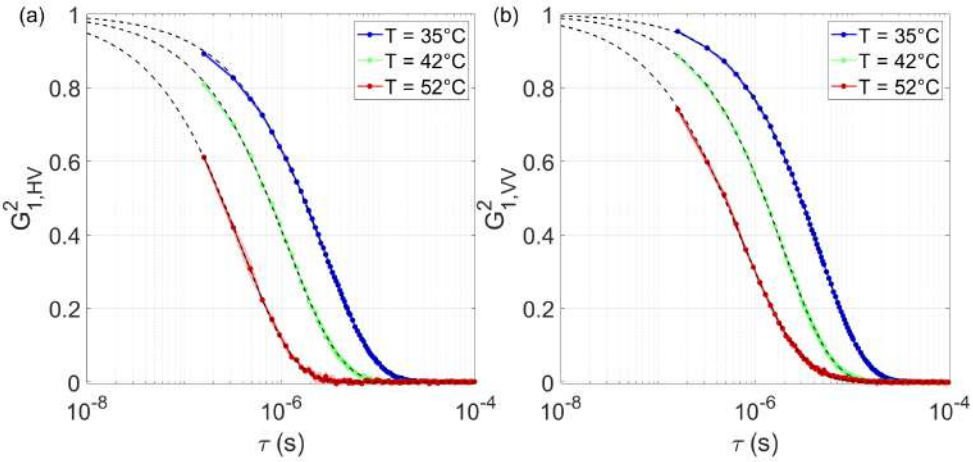

Figure 5.5: Squared field correlation functions $G^2_{1,HV}$ in panel(a) and $G^2_{1,VV}$ in panel(b) at $\theta = 90°$ at three different $T = 35°$C (blue dots), 42°C(green dots) and 52°C(red dots) for DD DNA duplexes at $c = 300$ g/l. Experimental data are fitted with eq.5.12 and 5.13. Dashed area marks experimental uncertainty.

where ΔG is the energy of the activated process, k_B is the Boltzmann constant. The logarithm of eq.5.8:

$$\ln\left(\frac{D}{D_0}\right) = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}, \tag{5.9}$$

allows to calculate the enthalpy ΔH, i.e. the magnitude of the attractive interaction between DD duplexes, simply from the slope of the linear fit of $\ln(D/D_0)$ data (see section5.5.1). In particular, the fits provide ΔH_{rot} = -18 kcal/mol and ΔH_{self} = -9 kcal/mol for rotational and self diffusion, respectively. From comparison with literature[10, 11], stacking energy is expected to be $\approx$ 10 kcal/mol in accordance with experimental ΔH_{self}, whereas ΔH_{rot} is too high, as found in SLS experiments in section4.4. Differently, collective diffusion displays a not trivial behaviour in temperature, starting from $D_{coll}/D_0 \approx 1$ at high T, increasing up to ≈ 1.8 in the intermediate range of T and finally decreasing down to ≈ 0.9 approaching T_{IN}. Although, the whole behaviour is hardly explainable compared to the nonlinear interacting DNA duplexes, the average lower value of D_{coll}/D_0 and especially its decrease towards T_{IN}, are a clear signature of attractive interactions between DD molecules, that lower the collective diffusion coefficient[12].

DD at intermediate concentration

Let's see now what happens diluting the solution down to c= 200g/l. Data were analyzed up to $T = 40°$, because at higher temperatures the intensity

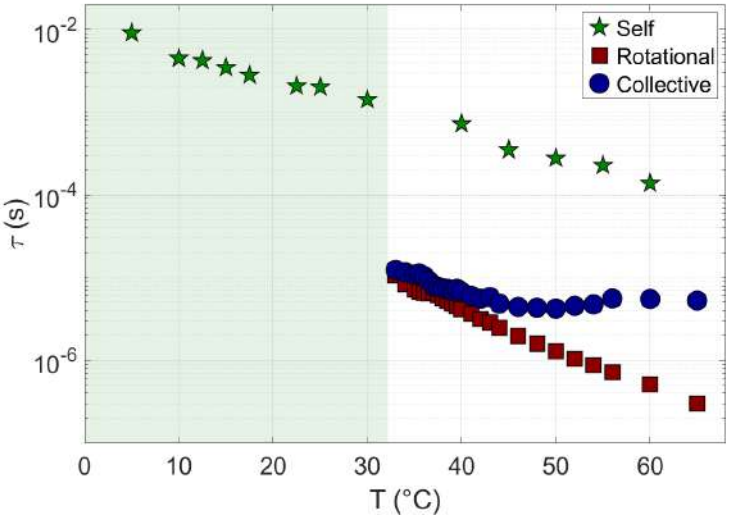

Figure 5.6: Decay times calculated from DLS and FRAP experiments for DD at c = 300 g/l. Symbols: green stars are calculated from diffusion coefficient of FRAP experiments; blue circles and green squares are rotational and collective decay times respectively, calculated from the fit of experimental data with eq.5.12 and 5.13.

field correlation functions in HV configuration were too noisy to provide reliable results. Fig.5.8 shows the squared field correlation functions. Similarly to the previous case, the $G^2_{1,HV}$ data are well fitted by a stretched exponential decay time due to rotational diffusion with $\alpha = 0.8$. Indeed, such decay time is independent on scattering angle θ and accordingly differs 2-3 order of magnitudes from the decay time corresponding to self diffusion. Even at lower concentration, at high T, the τ_{rot} is similar to DD and FT at high c ones. In Fig.5.9, the $G^2_{1,VV}$ data clearly show slower decay times compared to $G^2_{1,HV}$ over the all range of temperature explored. In particular, as shown in Fig.5.9, at low T the τ_{self} do not collapse on the τ_{coll} as for DD at c=300g/l. Such difference could be due to LC phase, that for DD at c=200g/l does not occur. Adimensional rotational and self diffusion exhibit an activated process. Fits according to eq.5.9 provides ΔH_{self} = -8kcal/mol and ΔH_{rot} = -8kcal/mol. So, the energy of activation related to rotation halved compared to ΔH at c = 300 g/l, whereas the ΔH related to self diffusion is unaffected by the change in concentration. Adimensional collective diffusion, shown in Fig.5.10, has an average value D_{coll}/D_0=0.9, confirming that attractive interaction reduce collective diffusion compared to the diffusion of repulsive molecules as FT duplexes at equal concentration.

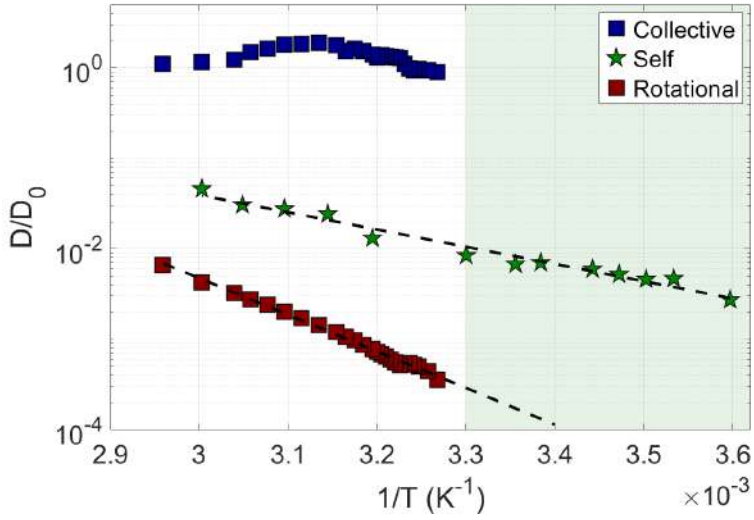

Figure 5.7: Adimensional diffusion coefficients D/D_0 vs. $1/T$ for DD DNA duplexes at c = 300g/l. Symbols: stars correspond to self diffusion coefficients calculated in FRAP experiments; squares and circles are rotational and collective diffusion coefficients, respectively. Black dashed lines show the exponential fit with eq.5.9, providing ΔH_{self}= -9kcal/mol and ΔH_{rot}-18 kcal/mol. In the green area the DNA solution is in the LC phase.

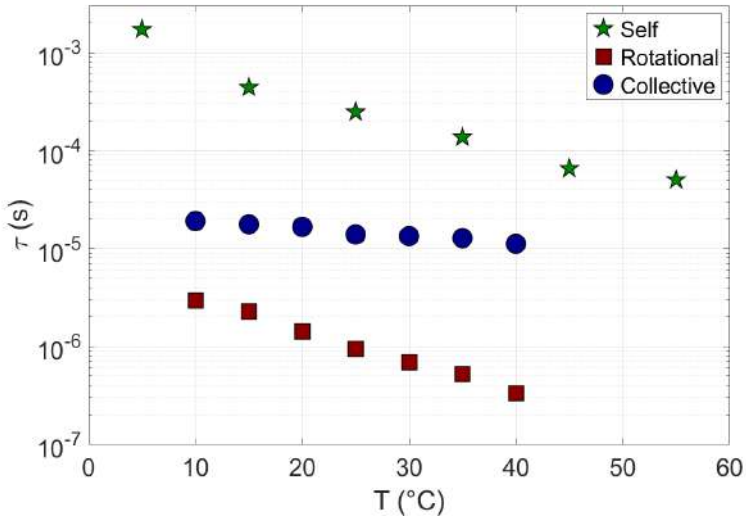

Figure 5.9: Decay times calculated from DLS and FRAP experiments for DD at c = 200g/l. Symbols: green stars are calculated from diffusion coefficient of FRAP experiments; blue circles and green squares are rotational and collective decay times respectively, calculated from the fit of experimental data with eq.5.12 and 5.13.

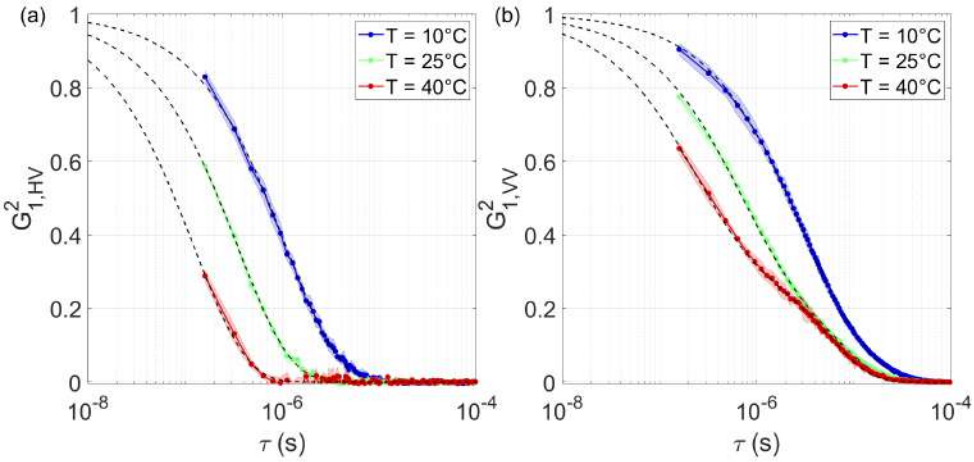

Figure 5.8: Squared field correlation functions $G^2_{1,HV}$ in panel(a) and $G^2_{1,VV}$ in panel(b) at $\theta = 90°$ at three different $T = 35°$C (blue dots), 42°C(green dots) and 52°C(red dots) for DD DNA duplexes at $c = 200$ g/l. Experimental data are fitted with eq.5.12 and 5.13. Shaded area marks experimental uncertainty.

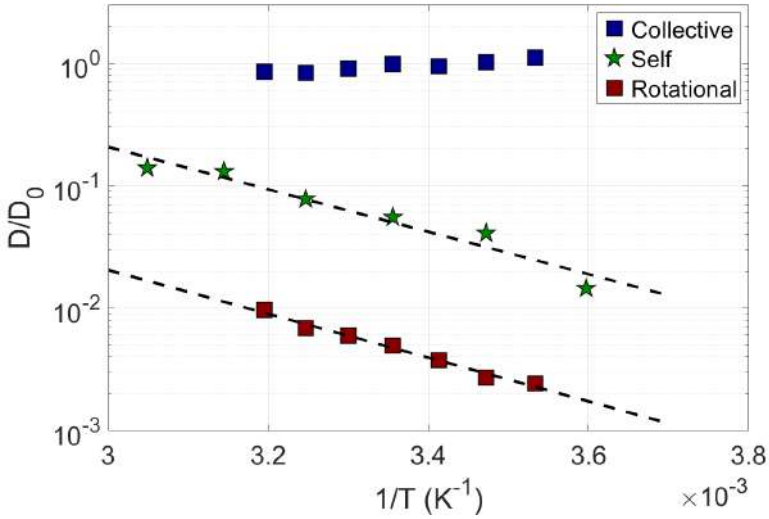

Figure 5.10: Adimensional diffusion coefficients D/D_0 vs. $1/T$ for DD DNA duplexes at $c = 200$g/l. Symbols: stars correspond to self diffusion coefficients calculated in FRAP experiments; squares and circles are rotational and collective diffusion coefficients from DLS experiments, respectively. Black dashed lines show the exponential fit with eq.5.9, providing ΔH_{self}= -8kcal/mol and ΔH_{rot}= -8 kcal/mol.

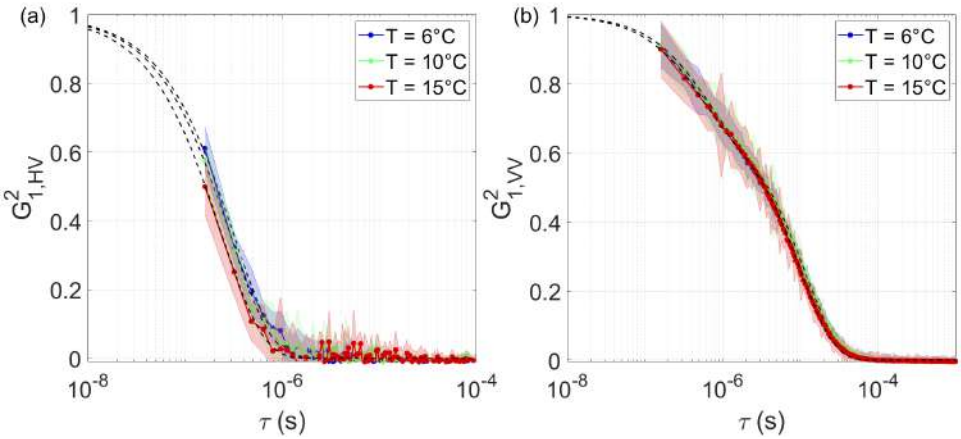

Figure 5.11: Squared field correlation functions $G^2_{1,HV}$ in panel(a) and $G^2_{1,VV}$ in panel(b) at $\theta = 90°$ at three different $T = 6°$C (blue dots), 10°C(green dots) and 15°C(red dots) for DD DNA duplexes at $c = 100$ g/l. Experimental data are fitted with eq.5.12 and eq.5.13. Shaded area marks experimental uncertainty.

DD at low concentration

It is interesting to investigate DD solution at c=100g/l to understand to what extent the aggregation process is further reduced. Data were analyzed up to $T = 15°$, because at higher temperatures the intensity field correlation functions in HV configuration were too noisy to provide reliable results. Fig.5.11 shows squared field correlation functions at three different temperatures. $G^2_{1,HV}$ data are well fitted by a stretched exponential decay time due to rotational diffusion with $\alpha = 1$, in accordance with a monodisperse solution. Such decay time is independent on scattering angle θ and accordingly differs almost 1 order of magnitude from the decay time corresponding to self diffusion at low T, as shown in Fig.5.12. In Fig.5.11(b), $G^2_{1,VV}$ data clearly show a slower decay time that differ 2 orders of magnitude from rotational decay times of $G^2_{1,HV}$. The behaviour over T of τ_{rot} and τ_{coll} shown in Fig.5.12 are very similar to FT duplexes ones at high c. However, at c=100g/l the τ_{rot} for FT duplexes are likely to be even shorter than DD ones. Indeed, as discussed in section5.3.1, the contrast of orientational decay is barely observable at c=200g/l. So, attractive interaction slows down rotational diffusion, whereas repulsive interaction makes it faster. Fig.5.13 shows that adimensional collective diffusion coefficient doesn't depend on T and it has an average value $D_{coll}/D_0 = 0.94 \pm 0.05$, similarly to DD at $c = 200$g/l. On the other hand, the self and rotational diffusion coefficients show an activated process whose fits according to eq.5.9 provides ΔH_{self} = -5kcal/mol and ΔH_{rot} = -1kcal/mol. Such strong reduction in ΔH_{rot} suggest that at such low concentration DD duplexes are almost free to rotate with a low energetic barrier to be overcome.

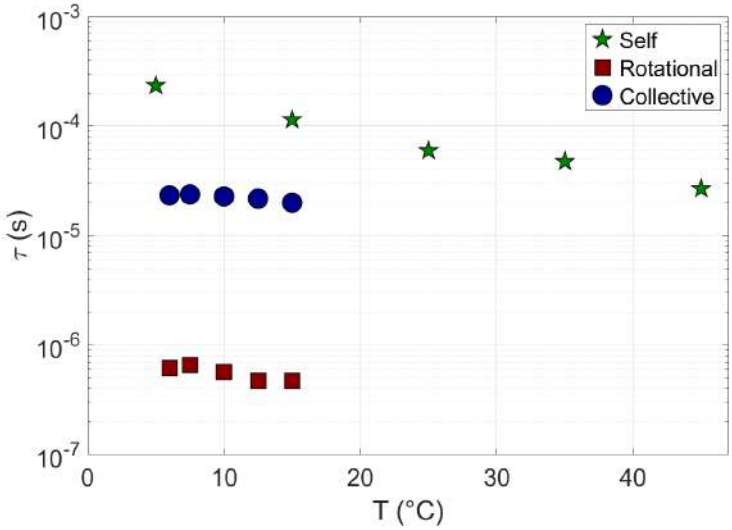

Figure 5.12: Decay times calculated from DLS and FRAP experiments for DD at c = 100 g/l. Symbols: green stars are calculated from diffusion coefficient of FRAP experiments; blue circles and green squares are rotational and collective decay times respectively, calculated from the fit of experimental data with eq.5.12 and eq.5.13.

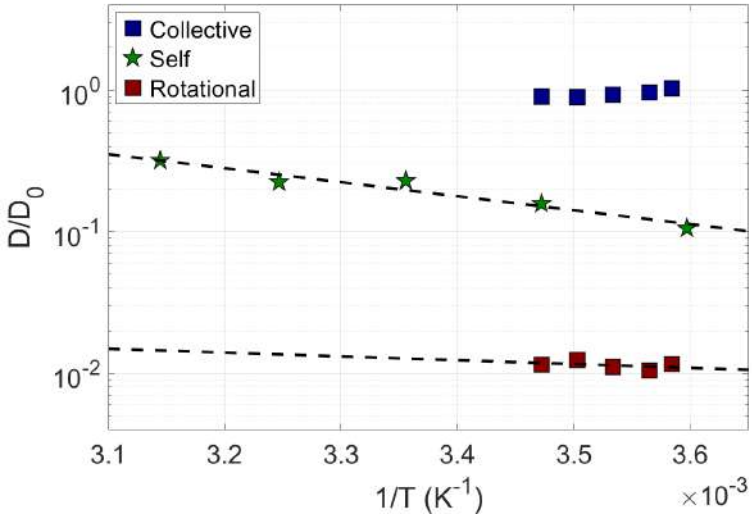

Figure 5.13: Adimensional diffusion coefficients D/D_0 vs. $1/T$ for DD DNA duplexes at c = 100g/l. Symbols: stars correspond to self diffusion coefficients calculated in FRAP experiments; squares and circles are rotational and collective diffusion coefficients from DLS experiments, respectively. Black dashed lines show the exponential fit with eq.5.9, providing ΔH_{self}= -5 and ΔH_{rot} -1 kcal/mol.

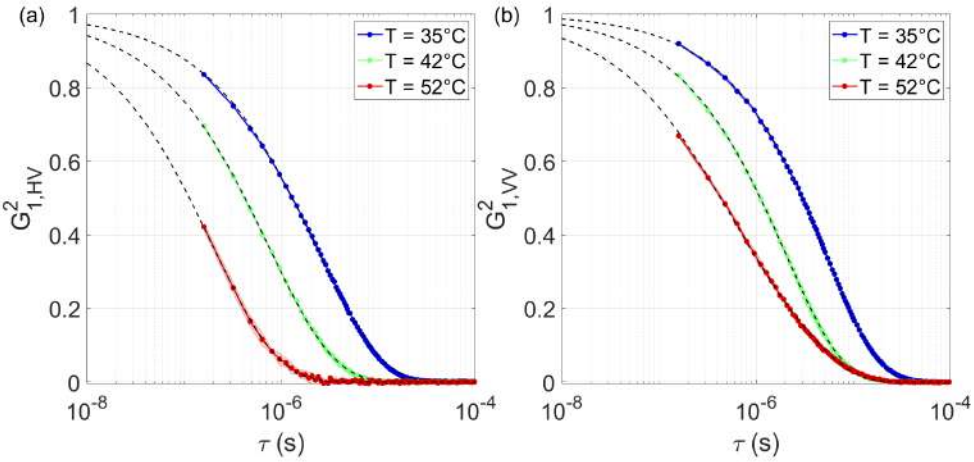

Figure 5.14: Squared field correlation functions $G^2_{1,HV}$ in panel(a) and $G^2_{1,VV}$ in panel(b) at $\theta = 90°$ at three different $T = 6°$C (blue dots), 10°C(green dots) and 15°C(red dots) for sDD DNA duplexes at $c = 300$ g/l. Experimental data are fitted with eq.5.12 and 5.13.

sDD at high concentration

Let's see what happens for sDD molecules whose attractive interaction is stronger compared to DD molecules and whose phase diagrams are described in[13]. Fig.5.14 shows experimental squared field correlation functions for sDD DNA duplexes at $c = 300$ g/l and at three different temperatures. $G^2_{1,HV}$ are well fitted with a stretched exponential decay with the stretched exponent $\alpha = 0.65$ (see eq.5.12). Such low value indicates a higher degree of polidispersity comapared to DD solution at the same c, whose α is equal to 0.8. The fast single decay time, which differs 2-3 order of magnitude from self decay time (see Fig.5.15) and its independence on the scattering angle θ confirm it to be due to rotational diffusion in accordance with theoretical model described by eq.5.3. The dependence on T is very similar to the one observed for DD duplexes at $c = 300$ g/l. This behaviour is likely to be due to the attractive interaction and growing cluster that hinder rotational diffusion more and more approaching the T_{IN}. The $G^2_{1,VV}$ in Fig.5.15(b) at T=35° and T=42° show a very similar decay to the corresponding $G^2_{1,HV}$, whereas at T=52° a slower decay time appears clearly. Fig.5.15 nicely displays the divergence of rotational and collective decay times at increasing T, similarly to DD at $c = 300$ g/l (see Fig.5.6). Fig.5.7 shows the adimensional diffusion coefficients. Collective diffusion is almost constant on T. At $T = 40$ °C, it starts decreasing down to $D_{coll}/D_0 = 0.4$ until it reaches T_{IN}. So, collective diffusion is even more strongly lowered by aggregation compared to the analogue solution of DD duplexes, where attractive interaction is weaker.

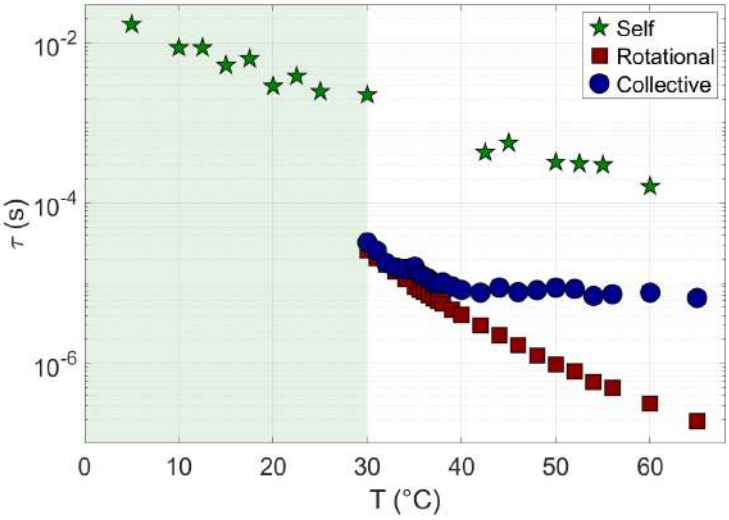

Figure 5.15: Decay times calculated from DLS and FRAP experiments for sDD at $c = 300$g/l. Symbols: green stars are calculated from diffusion coefficient of FRAP experiments; blue circles and green squares are rotational and collective decay times respectively, calculated from the fit of experimental data with eq.5.12 and 5.13.

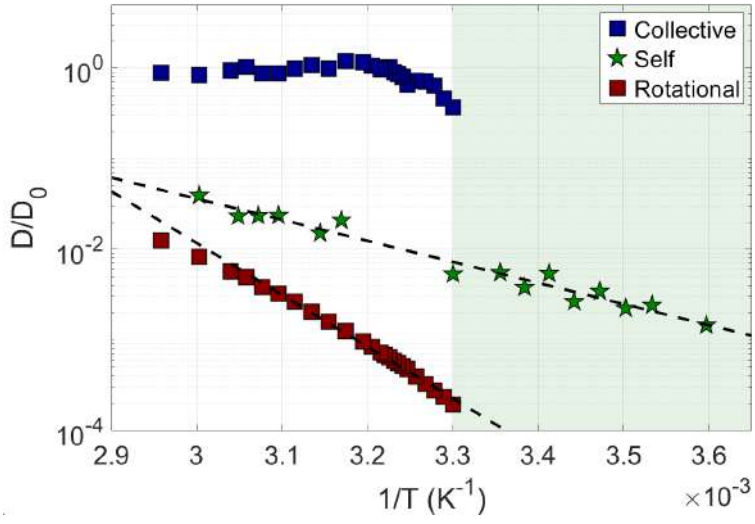

Figure 5.16: Adimensional diffusion coefficients D/D_0 vs. $1/T$ for sDD DNA duplexes at $c = 300$g/l. Symbols: stars correspond to self diffusion coefficients calculated in FRAP experiments; squares and circles are rotational and collective diffusion coefficients, respectively. Black dashed lines show the exponential fit with eq.5.9, providing ΔH_{self}= -11kcal/mol and ΔH_{rot} -28kcal/mol. In the green area the DNA solution is in the LC phase.

Such a difference is further confirmed by the activation energgy of self and rotational diffusion calculated from the fit according to eq.5.9. Indeed, ΔH_{self} = -11 kcal/mol, which is in accordance with the enthaply energy of a quartet CG/GC, $\Delta H = -10.6 kcal/mol$[14], whereas ΔH_{rot} = -26 kcal/mol. Even for sDD molecules, the activation energy for rotation is higher, more than double energy required to translate for a single molecule. Such a further increase could be due to a higher degree of orientational correlation between sDD duplexes compared to DD molecules.

5.4 Conclusions

In this section the main results of this work are discussed. Fig.5.17, nicely sums up the adimensional diffusion coefficients of all DNA solutions discussed so far.
The first result is the strong hampering of rotational diffusion which is surprisingly slightly dependent on the kind of interaction and concentration. Indeed, by looking at Fig.5.17, the adimensional D_{rot}/D_0 at the highest explored T for FT, DD and sDD molecules and at any concentration, have similar values. Although the origin of strong hampering of rotation is not completely understood, it can be described as macroscopic effect characterized by an increase of viscosity corresponding ≈40 times water viscosity(as shown for FT molecules in Fig.5.4).
The second result concerns the experimental evidence of activated process for the interacting molecules, unveiling a not trivial dependence of interaction energy on concentration. In particular, the experiments on the interacting DNA duplexes reveal an activated process for both rotational and self diffusion. If, on one hand the ΔH_{self} are in accordance with literature, on the other hand ΔH_{rot} are too high and strongly dependent on concentration. Such a behaviour could be explained in terms of pretransitional phenomena discussed in section4.4, where orientational correlation increases with concentration leading to a probable growth of the energetic barrier higher.
The third result is provided by the combination of D_{coll}/D_0 behaviour as a function of T and ΔH_{rot} that allow to distinguish both the quality and the strength of interaction among duplexes. Indeed, it has been shown that the collective diffusion of FT molecules is nicely described in terms of repulsive hard spheres, whereas the decrease of D_{coll}/D_0 for sDD and DD molecules approaching T_{IN} evidence an attractive interaction among duplexes. The strength of attractive interaction can be distinguished by looking at ΔH_{self} and ΔH_{rot} values, that are higher for sDD than for DD duplexes. In particular, ΔH_{self} = -11 kcal/mol and ΔH_{rot} = -26 kcal/mol for sDD molecules, versus ΔH_{self} = -8 kcal/mol and ΔH_{rot} = -18 kcal/mol for DD molecules.
The fourth result found by Dr. Todisco in FRAP experiments regards the non

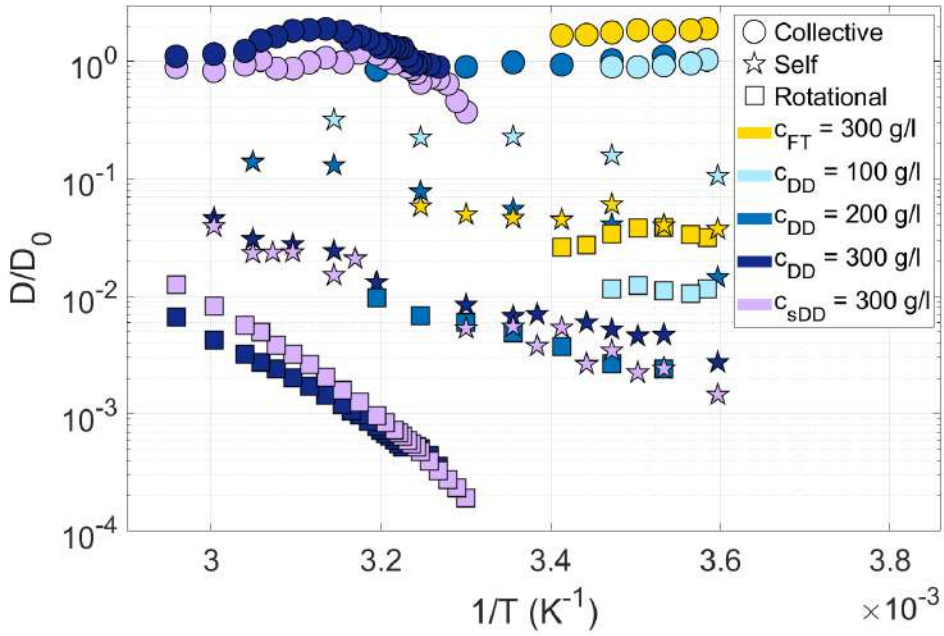

Figure 5.17: Summary plot of adimensional diffusion coefficients for FT, DD, sDD duplexes. Symbols: circles, stars and squares correspond to collective, self and rotational diffusion coefficients, respectively. Different colours correspond to different kind of DNA molecules and concentration, as specified in the legend.

discontinuity in difussion coefficient from isotropic to liquid crystal phase. A possible explanation for this phenomenon is that the slow-down in LC phase is compensated by the removal of excluded volume from short duplexes terminals.

5.5 Material and Methods

5.5.1 Analysis of intensity correlation functions

Fig.5.18 shows the steps in the analysis of both polarized and depolarized intensity correlation functions, $G_2(\tau)$, to get the squared field correlation functions, $|G_1(\tau)|^2$ for DD and sDD DNA solution. Firstly, the $G_{2,HV}(\tau)$ are fitted with (see Fig.5.18(a)):

$$y_{G_{2,HV}} = a_1 + a_2 e^{-2(\tau/a_3)^{\alpha} - 2D_T^{self} q^2 \tau}, \tag{5.10}$$

where a_i are free parameters, α is the stretch exponent and D_T^{self} is the self diffusion coefficient measured in FRAP experiments. The parameter a_3 corresponds to the τ_{rot}. Secondly, the $G_{2,VV}(t)$ are fitted with(see Fig.5.18(b)):

$$y_{G_{2,VV}} = b_1 + b_2\left(b_3 e^{-(\tau/\tau_{rot})^{\alpha}} + (1-b_3)e^{-\tau/b_4}\right)^2, \tag{5.11}$$

where b_i are free parameters, τ_{rot} is the rotational diffusion time calculated from the fit of $G_{2,HV}(t)$ data. Finally, the squared field correlation functions $|G_{1,HV}|^2(|G_{1,VV}|^2)$ are calculated from experimental $G_{2,HV}$ (G_{2VV}) subtracting the baselines, $a_1(b_1)$, and dividing by the contrast $a_2(b_2)$. Square field correlation functions in HV configuration are then fitted with(see Fig.5.18(c)):

$$y_{G^2_{1,HV}} = e^{-2(\tau/c_1)^{\alpha} - 2D_T^{self}q^2\tau}, \tag{5.12}$$

where c_1 is the only free parameter, whereas $|G_{1,VV}|^2$ are fitted with(see Fig.5.18(d)):

$$y_{G^2_{1,VV}} = \left(d_1 e^{-(\tau/\tau_{rot})^{\alpha}} + (1-d_1)e^{-\tau/d_2}\right)^2, \tag{5.13}$$

where d_i are free parameters, τ_{rot} is the rotational diffusion time equal to the parameter c_1 calculated from eq.5.12.

As mentioned in section5.3.1, for FT duplexes it was not possible to calculate $G_{2,HV}$ from scattered light in HV configuration. Consequently, experimental $G_{2,VV}$ were fitted with eq.5.11, where the rotational time, τ_{rot}, was set also as free parameter and α=1. The following steps in the data analysis to get the $|G_{1,VV}|^2$ are identical to those described above.

To calculate the effective viscosity η_{eff}, the $|G_{1,VV}|^2$ were also fitted globally by:

$$y_{G^2_{1,VV}} = \left(d_1 e^{-(6D_{rot}/\gamma\tau)} + (1-d_1)e^{-D_T^{coll}q^2\tau}\right)^2, \tag{5.14}$$

where D_{rot} is calculated according to eq.2.12, D_T^{coll} is calculated from the fit with eq.5.13, d_1 is a free parameter for each curve, whereas γ is a global free parameter. The parameter γ is the factor multiplying the water viscosity, so that such a global fit provides the effective viscosity, $\eta_{eff} = \gamma\eta_{water}$ that molecules feel rotating.

Activated process

The logarithm of adimensional self and rotational diffusion coefficients that show an activated process as a function of $1/T$ were fitted with:

$$y = a_1 + a_2\frac{1}{T} \tag{5.15}$$

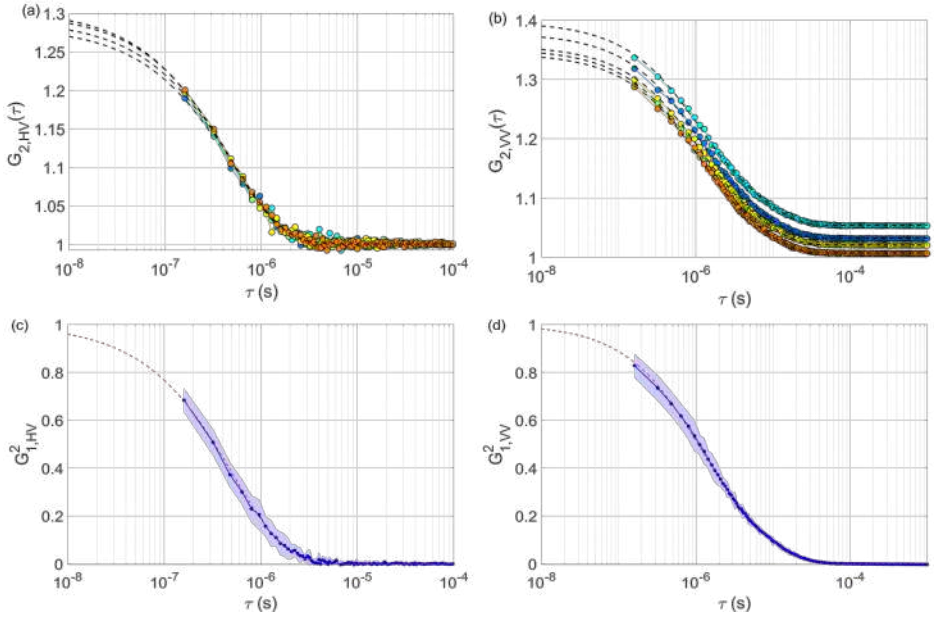

Figure 5.18: Steps in the analysis of depolarized and polarized intensity correlation functions. (a) Raw data of intensity correlation functions in depolarized configuration, fitted with eq.5.10. (b) Raw data of intensity correlation functions in polarized configuration, fitted with eq.5.11(dashed lines). (c) Average squared filed correlation function in depolarized configuration obtained after normalizing and averaging data in panel(a). The average $G^2_{1,HV}$ is fitted with eq.5.12(red dashed line). (d)Average squared filed correlation function in depolarized configuration obtained after normalizing and averaging data in panel(b). The average $G^2_{1,VV}$ is fitted with eq.5.13(red dashed line). Shaded light blue area marks the experimental uncertainty.

where a_i are free parameters. From comparison with eq.5.9, the activation energy was calculated as:

$$\Delta H = \frac{a_2 R}{4.2\ 10^{-3}}, \tag{5.16}$$

where R is the gas constant, and the value at the denominator convert the energy in kcal/mol.

5.5.2 DNA sample preparation

FT, DD and sDD were firstly diluted and dialysed against $25mM$ NaCl, thus fixing the molar ratio between nucleotides and ions in the preparation batches. Hilgenberg round capillaries of 2.4 mm inner diameter and solution volume $\approx 30\mu l$ were prepared at $c_{DNA} = 300 - 200 - 100 \quad mg/ml$ for DD, at $c = 300 - 200 \quad mg/ml$ for FT and at $c = 300 \quad mg/ml$ for sDD. All DNA solutions have ionic strength $c_{NaCl} = 2M$.

Bibliography

[1] Michi Nakata, Giuliano Zanchetta, Brandon D Chapman, Christopher D Jones, Julie O Cross, Ronald Pindak, Tommaso Bellini, and Noel A Clark. End-to-end stacking and liquid crystal condensation of 6–to 20–base pair dna duplexes. *Science*, 318(5854):1276–1279, 2007.

[2] Gregory P Smith, Tommaso P Fraccia, Marco Todisco, Giuliano Zanchetta, Chenhui Zhu, Emily Hayden, Tommaso Bellini, and Noel A Clark. Backbone-free duplex-stacked monomer nucleic acids exhibiting watson–crick selectivity. *Proceedings of the National Academy of Sciences*, 115(33):E7658–E7664, 2018.

[3] R Piazza, J Stavans, T Bellini, D Lenti, M Visca, and V Degiorgio. Light-scattering experiment on anisotropic spherical particles. *Trends in Colloid and Interface Science IV*, pages 89–94, 1990.

[4] Roberto Piazza and Vittorio Degiorgio. Rotational and translational self-diffusion coefficients of interacting brownian spheres. *Journal of Physics: Condensed Matter*, 5(34B):B173, 1993.

[5] Chris R Calladine and Horace Drew. *Understanding DNA: the molecule and how it works*. Academic press, 1997.

[6] W Richard Bowen and Anne Mongruel. Calculation of the collective diffusion coefficient of electrostatically stabilised colloidal particles. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 138(2-3):161–172, 1998.

[7] A Van Blaaderen, J Peetermans, G Maret, and JKG Dhont. Long-time self-diffusion of spherical colloidal particles measured with fluorescence recovery after photobleaching. *The Journal of chemical physics*, 96(6):4591–4603, 1992.

[8] B Cichocki, ML Ekiel-Jeżewska, and E Wajnryb. Lubrication corrections for three-particle contribution to short-time self-diffusion coefficients in colloidal dispersions. *The Journal of chemical physics*, 111(7):3265–3273, 1999.

[9] Marco Todisco, Gregory P Smith, and Tommaso P Fraccia. Liquid crystal ordering of dna dickerson dodecamer duplexes with different 5'-phosphate terminations. *Molecular Crystals and Liquid Crystals*, 683(1):69–80, 2019.

[10] Cristiano De Michele, Lorenzo Rovigatti, Tommaso Bellini, and Francesco Sciortino. Self-assembly of short dna duplexes: from a coarse-grained model to experiments through a theoretical link. *Soft Matter*, 8(32):8388–8398, 2012.

[11] Daniel Svozil, Pavel Hobza, and Jiri Sponer. Comparison of intrinsic stacking energies of ten unique dinucleotide steps in a-rna and b-dna duplexes. can we determine correct order of stability by quantum-chemical calculations? *The Journal of Physical Chemistry B*, 114(2):1191–1203, 2010.

[12] Mario Corti and Vittorio Degiorgio. Quasi-elastic light scattering study of intermicellar interactions in aqueous sodium dodecyl sulfate solutions. *The Journal of Physical Chemistry*, 85(6):711–717, 1981.

[13] Marina Rossi, Giuliano Zanchetta, Sven Klussmann, Noel A Clark, and Tommaso Bellini. Propagation of chirality in mixtures of natural and enantiomeric dna oligomers. *Physical review letters*, 110(10):107801, 2013.

[14] John SantaLucia Jr and Donald Hicks. The thermodynamics of dna structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440, 2004.

Chapter 6

DNA-polycations coacervates

6.1 Introduction

In living cells there are various membraneless organells with inner structures that are capable to support biological activities[1] and which are also involved in diseases where protein aggregation occurs in. These are just some of examples that, in the last decade, stimulated the interest in understanding the crucial features that lead to biomelocular condensation[2].
The onset of these organelles is due to a phase transition in liquid-liquid phase separated systems (LLPS). This phenomena is also called coacervation[3] and it occurs in mixtures of opposite charged polyelectrolytes. Indeed, because of electrostatic interactions and entropic gain[4], a concentrated but still liquid[5, 6] phase of polyelectrolytes develops (coacervate) in a bulk phase where polyelectrolytes are diluted. The main differences between coacervates and precipitates are the water content, being lower for solid aggregates, and the state of matter, which is amorphous for precipitates and liquid for coacervates.
The two main ingredients that determine coacervation are salt concentration and the total charge of polyelectrolytes. In particular, the higher the salt concentration and the more the polyelectrolytes are screened, thus strongly reducing their electrostatic interactions[7]. So, it exists a critical salt concentration above which coacervates disassemble in a isotropic uniphase. On the other hand, an high total charge concentration $Q = N + P$, where N and P indicate the total negative and positive charge concentration respectively, enhances coacervation[6, 4].
Here, the negative polyelectrolite of interest is the DNA molecule, where negative charge is due to phosphate groups. The interaction of single stranded DNA (ssDNA) and double stranded DNA (dsDNA) with positive polyelectrolytes is different, favouring coacervation for ssDNA and precipitation for dsDNA. Both the higher charge and lower flexibility of dsDNA make coacervation process more difficult leading to the development of precipitates[8]. Despite these drawbacks, surprisingly it has been shown that liquid crystal ordering can develop in coacervates of short dsDNA 22 and 12 nucleobases long with poly-L-lysine (PLL)[9, 6]. Such a phenomena is called liquid-liquid crystal phase separation (LLCPS). In particular, in[6] it has been shown that

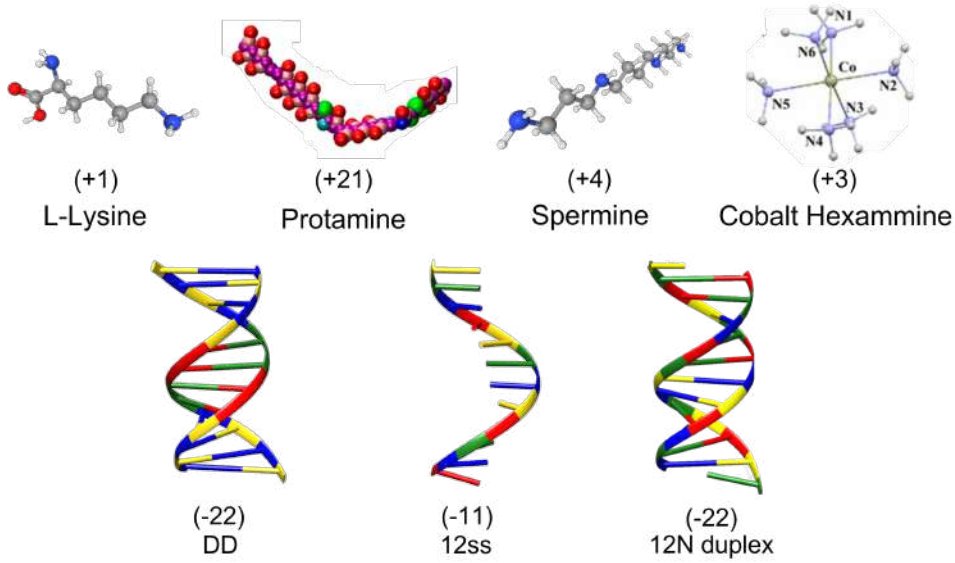

Figure 6.1: Overview of polycations and DNA molecules used to investigate liquid-liquid phase separation. Polycations are not in scale. L-Lysine is the repeated unit in poly-l-lysine, whose average total charge is +240. Protamine: spheres represent different amino acids, in particular 21 red balls are the positive arginine amino acids (adapted from[10]). Cobalt Hexammine adapted from[11] has positive total charge +3.DNA chain of length L has L-1 negative charges. 12N duplex has been represented as the most probable duplex at low T with α_s=1, as shown in Fig.3.9.

	POLY-L-LYSINE	PROTAMINE	SPERMINE	COBALT-HEX
DD	✓	✓	✓	✓
12A	✓	✓	-	-
12N	✓	✓	-	-

Figure 6.2: The ticks mark the combination of DNA and polycations experimentally investigated.

mixtures of blunt-end reverse DD (rDD:5'-GCGCTTAAGCGC-3') with PLL exhibits both isotropic phase and all the LC phases as the N^* and columnar.

Inspired from such outstanding work, I investigated LLPS in aqueous solutions of 12 nucleobases long DNA oligomers and positive charged electrolytes. Fig.6.1 shows all DNA molecules and electrolytes used, whereas Fig.6.2 displays all the combination of DNA and polycations explored here among the polyelectrolytes shown in Fig.6.1.

The goal of this preliminary and qualitative work is to explore the phase behaviour of each combination reported in Fig.6.2 by observing the mesophases as a function of temperature and salt concentration. All the experiments were performed in charged-balanced samples. Initially, the solutions are prepared at room temperature. Successively, the solutions are annealed by heating the samples up to 90°C and then by cooling with a ramp rate $\Delta T/t \leq 5°/\text{min}$. In the following, for each DNA -polycations combination, only the cooling phase diagrams are reported.

6.2 Materials and Methods

All the polycations were purchased from Sigma-Aldrich. In the following I report the specifications with corresponding catalog numbers: PLL (poly-L-lysine hydrobromide from 30to 70 kDa, P2636), protamine (Protamine sulfate salt from salmon, P4020), cobalt hexamine (Hexaamminecobalt(III) chloride, 481521), spermine (Spermine tetrahydrochloride, S1141). DNA single strand 12ss (5'-CGACCTGTCACT-3') was purchased by IDT. All DNA sequences were dialyzed against 1.5 L of 25 mM NaCl for 48 hours. The dialysis buffer was replaced after 8 and 24 hours. Successively, the DNA solution were liophilyzed and diluted in Milli-Q water in stock solution of c_{DNA} ≈30g/l. The concentration was measured with UV-absorbance with UV-Vis 300 spectrophotometer as explained in Chapter3.

Sample preparation

All DNA stock solutions were prepared at c = 100 mM of total charge, where the negative charges are from phosphate groups (P). All cationic stock solutions were prepared at c = 100 mM, but protamine at c = 85 mM of total positive charge. Two different NaCl stock solutions were prepared at c_{NaCl} = 4 M and c_{NaCl} = 500 mM. Buffer stock solution of Hepes was prepared at c_{Hepes} = 250 mM and pH=7. These stock solutions were used to prepare samples of total volume V_{tot}= 10 μl in 0.2 mL plastic centrifuge tube from Eppendorf, with total charge concentration Q=20mM (or Q=18.5mM for solutions with protamine), salt concentration in the range of 0-1.2M, and c_{Hepes}=25mM. The samples were prepared by adding progressively as the following: Milli-Q water, Hepes, NaCl, cationic compound and DNA. The solutions were prepared at room temperature.
Plastic plates (Nunc 96-Well Cell Culture Plates) were firstly cut to be hosted in temperature-controlled chamber on microscope stage. Then the wells were passivated by covering them with bovine serum albumin (BSA) solution, whose mass fraction is 3.5%, for 15 min. Then, the well are rinsed and dried. Solutions of DNA and cationic compounds were located on the bottom. They were covered with mineral oil (M5904, Sigma-Aldrich) and a

plastic PCR film was applied on the plastic plate to avoid evaporation. The well with glass bottom was not passivated with BSA, and it was covered with mineral oil and and a glass cover slip.

Experimental setup

Plastic plates were observed with Nikon TE200 microscope and PTOM images acquired with Nikon DS-5M CCD camera. To explore as much as possible mixtures simultaneously, aqueous solution was inserted in resized plastic wells to be hosted in the temperature controlled chamber on stage microscope. For this reason, almost all PTOM reported images are affected by birefringence of plastic plates that makes difficult to identify the kind of LC phase. Only for 12N-PLL at one specific salt concentration, the solution was inserted in a well with bottom glass slide that allows to take standard PTOM images and recognition of LC phase.

6.3 Phase Diagrams

In the following, for each polyelectrolytes combination described in Fig.6.2, the phase diagrams and corresponding observations are reported.

6.3.1 DD and PLL

The combination here reported is the aqueous solution of DD DNA duplexes and PLL with Q=20mM (N=P), as investigated in[6] where rDD has been used on behalf of DD. DD duplexes support LC ordering at concentration c>240-300 g/l[12], and each DNA chain has one negative charge per phosphate, so L-1 (L is the length of DNA chain) negative charges (11-). Instead, PLL, whose molecular weight MW = 30 -70 kDa corresponds to an average number of 240 lysine per chain, at pH = 7, each Lys ϵ-amino group has a positive charge, so that the average total charge per chain is 240+.
Initially, all the samples corresponding to c_{NaCl} ≤450mM show precipitates birefringent structures as reported in[6], whereas at 450≤ c_{NaCl} ≤650mM coacervate isotropic droplets form. At $c_{NaCl} \geq 650$ mM no coacervates are detected. After annealing process, the insurgence of LC phase in a wide range of c_{NaCl} was observed. As shown in Fig.6.3, LC phases were detected up to c_{NaCl}=450mM, whereas isotropic coacervates occur in the narrow range of salt concentration 450<c_{NaCl}<650mM. At $c_{NaCl} = 650$ mM no LLPS is detected. Although the phase diagrams in Fig.6.3 is shifted of ≈ 200mM toward higher c_{NaCl} compared to the one reported in[6] at Q=20mM, the temperature behaviour is almost in accordance with. Such salt concentration difference could be partially due to a different NaCl content in DNA

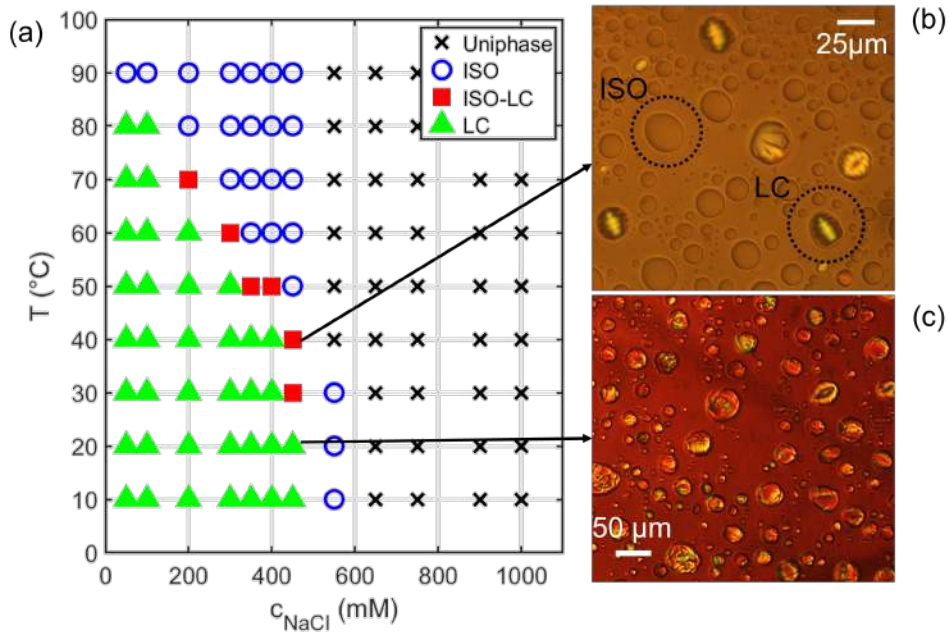

Figure 6.3: Panel(a). c_{NaCl}-T phase diagram for aqueous solution of DD and PLL, obtained upon cooling from 90°C and PTOM images corresponding to salt concentration and T indicated by arrows. Panel(b) shows coexistence of coacervates in isotropic and LC phase. Panel(c). By decreasing the temperature, in all coacervates LC order occurs.

stock solutions. Moreover, the phase behaviour of DD-PLL solutions is here extended to lower c_{NaCl} where LC phases are observed to be stable up to T=80°C for $c_{NaCl\leq}$100mM. The latter result was really surprisingly. Indeed, even if the DNA concentration was as high as 500-700 g/l, at T=80°C half of DD duplexes are predicted to be melted[13], and consequently LC ordering should be suppressed. Thus, it's likely that DNA-PLL interaction increase energetic stability of DD duplexes and consequently of LC ordering.

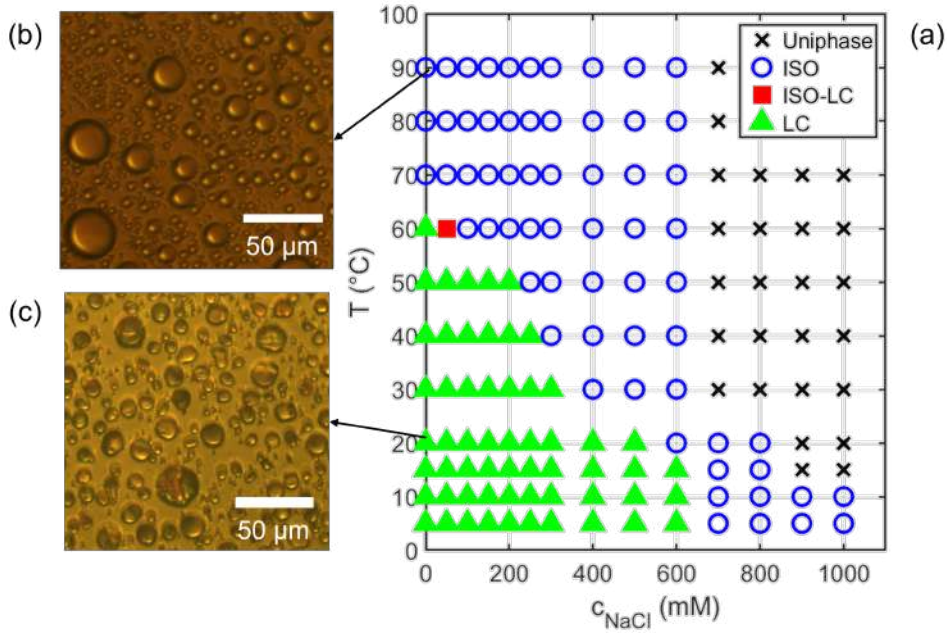

Figure 6.4: Panel(a).c_{NaCl}-T phase diagram for aqueous solution of DD and protamine, obtained upon cooling from 90°C and PTOM images corresponding to salt concentration and T indicated by arrows. Panel(b). Isotropic coacervates at T=90°C. Panel(c). Example of LC onset in coacervates.

6.3.2 DD and Protamine

The combination here reported is the aqueous solution of DD DNA duplexes and protamine. Protamine is a protein with MW = 5.1 kDa composed by 30 amino acids whose 21 over 30 are positively charged (+21). For DNA protamine solution the neutral charge is achieved by ratio between positive protamine and negative DNA charges $R_{+/-} = 0.85$[14]. Steric and electrostatic interactions between protamine and DNA duplexes 50 base pair long have been computationally investigated in[10], unveiling the formation of protein bridges that favour attraction between DNA molecules leading to the formation of large bundle, where positional hexagonal order occurs. Such a feature has been also observed in CryoTEM experiments with DNA molecule 146 nucleobases long[10]. To the best of our knowledge no experimental evidences of LC ordering in coacervates of protamine and DNA duplexes with length < 146 base pair have been reported. The comparison between the phase diagrams of DD-protamine shown in Fig.6.4 and DD-PLL shown in Fig.6.3, surprisingly reveals on one hand that DD-protamine interaction extends LC ordering in even a wider range of salt concentration, up to c_{NaCl}=600mM, and on the other hand it reduces thermal stability of LC phases for all $c_{NaCl} \leq$=400mM

of $\Delta T \approx$ 10-20°C. Oppositely, isotropic coacervates occur in a wider range of salt concentration, up to c_{NaCl} = 1M and they are more stable over T.

6.3.3 DD with Spermine

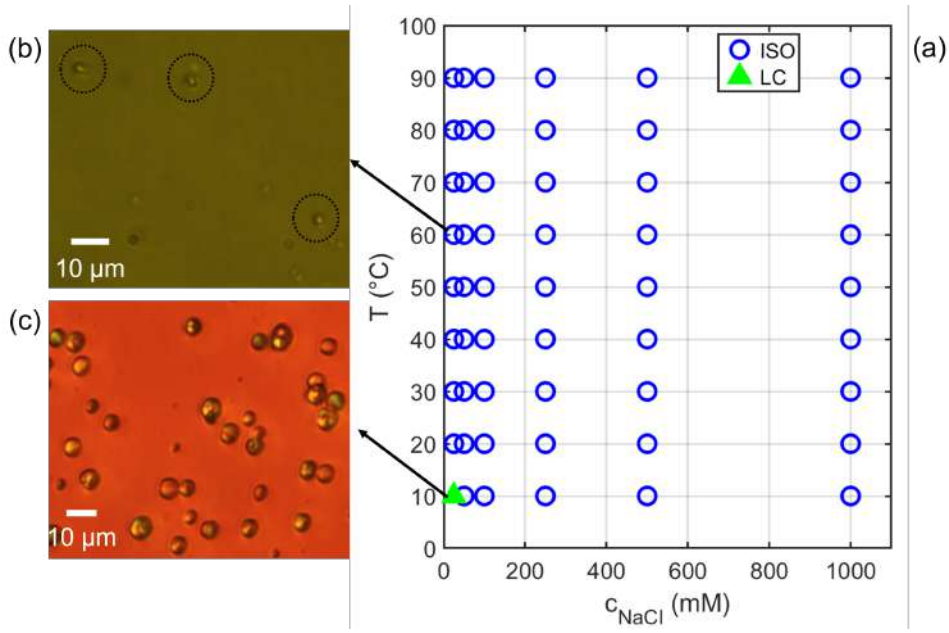

Figure 6.5: Panel(a). c_{NaCl}-T phase diagram for aqueous solution of DD and spermine, obtained upon cooling from 90°C and PTOM images corresponding to salt concentration and T indicated by arrows. Panel(b). Isotropic coacervates Panel(c).

The combination here reported is the aqueous solution of DD DNA duplexes and spermine with Q=20mM (N=P). Spermine is a polyamine that at pH=7 all its four nitrogen atoms are protonated (+4). For all the salt concentration reported in Fig.6.5 has been observed phase separation. A characteristic feature of such isotropic coacervates is their size. As shown in Fig.6.5(b) they are barely visible even at the highest magnification. However, at c_{NaCl}=25 mM decreasing the temperature down to 10°C, the size of such droplets increase up to reveal the insurgence of LC order. In literature[15], solution of spermine and dsDNA 25 nucleobases long exhibits not well identifiable phase states.

6.3.4 DD with Cobalthexamine

The phase diagrams of aqueous solution of DD DNA duplexes and cobalthexamine with Q=20mM (N=P). Cobalthexamine is an inorganic cation with

three positive charges (3+), is identical to the phase diagram of DD and spermine with the exception of not LC phase. The size of isotropic coacervates are similar to the ones shown in Fig.6.5(b) and it's very hard to distinguish them from a uniphase.

6.3.5 12N and Poly-lysine

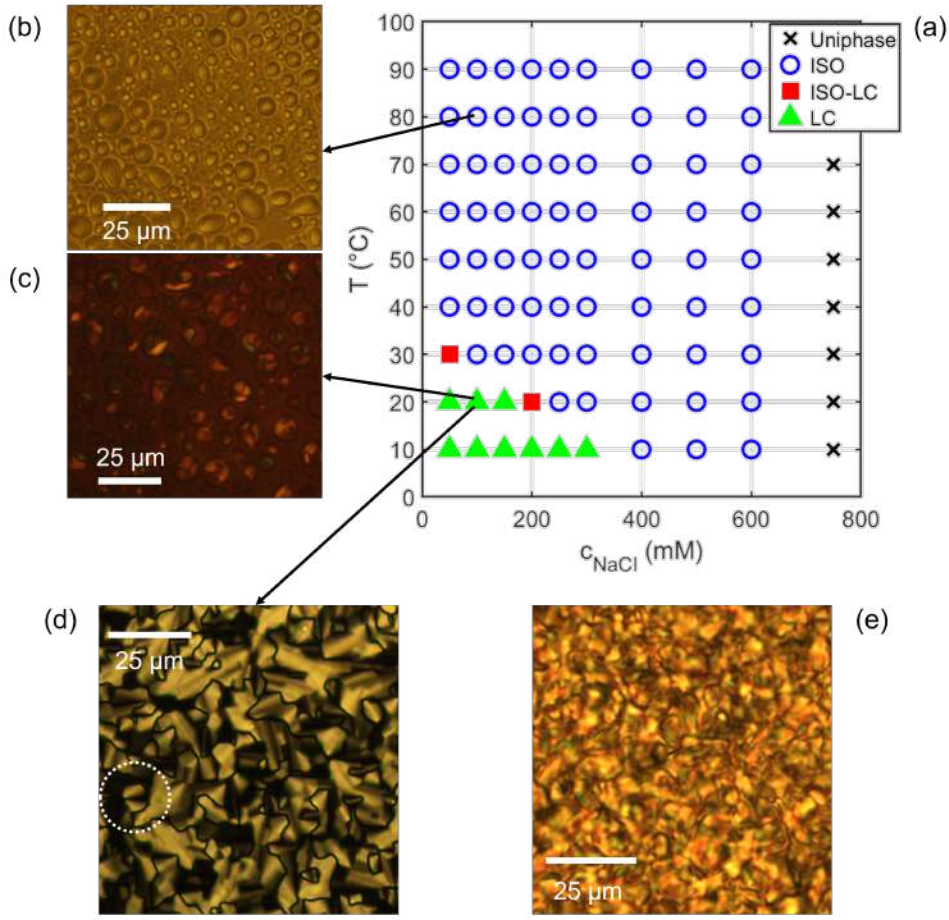

Figure 6.6: Panel(a)c_{NaCl}-T phase diagram for aqueous solution of 12N and PLL, obtained upon cooling from 90°C. Panel(b-c-d) are PTOM images corresponding to salt concentration and T indicated by arrows. PTOM images shown in(b-c) are taken in plastic plates, whereas the image shown in(d) is taken in glass plate. Dashed circle higlights focal conics, that is a characteristic feature of columnar phase. Panel(e) PTOM image of 12N solution at c = 910g/l that shows a different and hardly identifiable LC phase.

The combination here reported is the aqueous solution of 12N DNA duplexes and PLL with Q=20mM (N=P). Fig.6.3(a)and(c-d) show the onset of LC

ordering in 12N-PLL coacervates at $T \leq 20°C$ and $c_{NaCl} \leq 300mM$. To identify the LC phase, the solution at $c_{NaCl} = 100mM$ has been observed in well with bottom glass and standard PTOM images could have been taken. In this sample the liquid-liquid phase separation occurred in larger size coacervates and as shown in Fig.6.3(d) it reveals a columnar LC phase identified by focal conics (highlighted by dashed circle). Such a result was as surprising as it was unexpected. Indeed, as shown in Fig.6.3(e), I have experimentally proved that 12N DNA solution show LC ordering at really high concentration, $c_{12N} \approx 900g/l$, showing a LC phase hardly identifiable. And, in[6] the highest measured c_{DNA} in DD-PLL coacervates with columnar LC phase is c_{DNA}=600g/l. For these reasons, it's really surprising to observe LC ordering in 12N-PLL solutions and specifically a columnar LC phase. Such a outcome further confirms on one hand the results found in Chapter3, where at low temperature most duplexes have less than two terminal shift errors, (with higher probability to sustain LC ordering compared to duplexes with more than two pairing errors) and on the other hand it strengthens the hypothesis that PLL favours LC ordering.

6.3.6 12N and Protamine

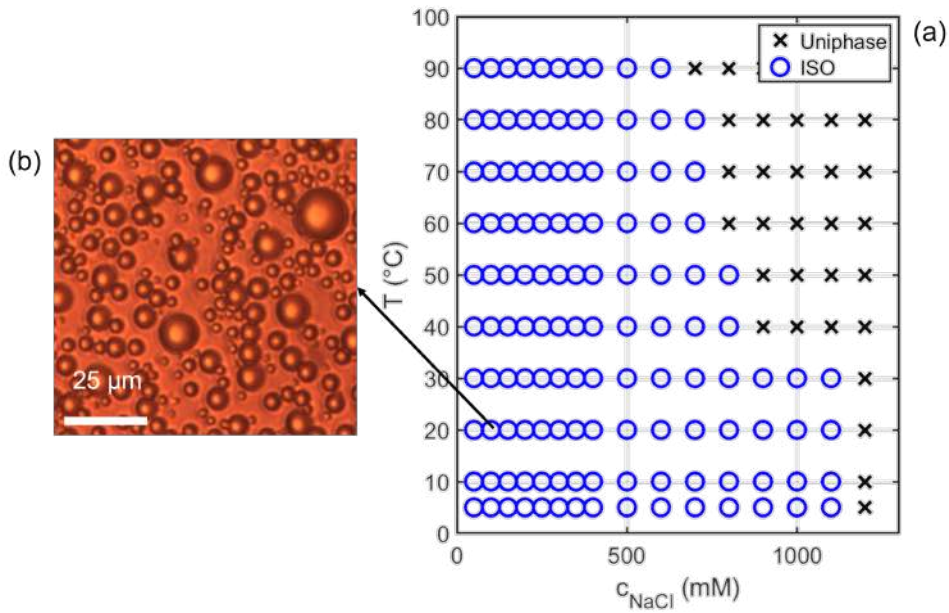

Figure 6.7: Panel(a).c_{NaCl}-T phase diagram for aqueous solution of 12N and protamine, obtained upon cooling from 90°C and PTOM image corresponding to salt concentration and T indicated by arrow. Panel(b). Example of isotropic coacervates.

The combination here reported is the aqueous solution of 12N DNA duplexes and protamine with Q=18.5mM ($R_{+/-}$=0.85). Fig.6.7 shows the corresponding phase diagram. Although no LC ordering is observed, isotropic coacervates occur until c_{NaCl}=1.1M. So, the effect of interaction between protamine and 12N duplexes increases the critical salt concentration compared to 12N-PLL as analogously observed in both DD-PLL and DD-protamine phase diagrams.

6.3.7 12ss and PLL

The combination here reported is the aqueous solution of 12ss DNA duplexes and PLL with Q=20mM. Solution of single strand DNA (10-88 nucleobases long) and PLL (10-100 amino acids long) have been shown to form isotropic coacervates in a wide range of salt concentration (0 -700mM)[8]. Indeed, the phase diagram in Fig.6.8 reports the formation of isotropic coacervates up to c_{NaCl}=450mM, above which coacervates disassemble.

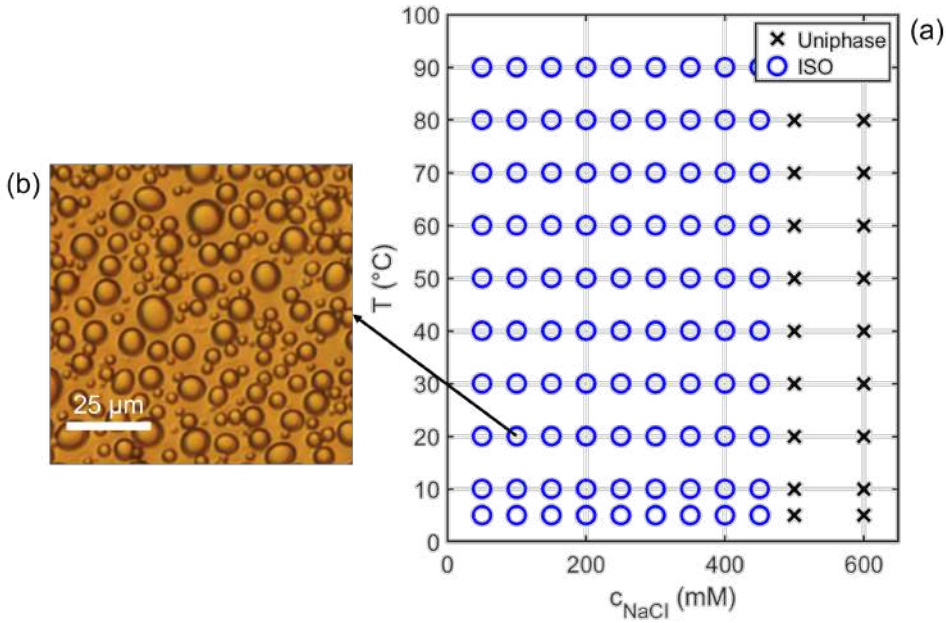

Figure 6.8: Pnael(a).c_{NaCl}-T phase diagram for aqueous solution of 12A and PLL, obtained upon cooling from 90°C and PTOM image corresponding to salt concentration and T indicated by arrow.Panel(b). Example of isotropic coacervates.

6.3.8 12ss and Protamine

The combination here reported is the aqueous solution of 12ss DNA duplexes and protamine with Q=18.5mM ($R_{+/-}$=0.85). Phase diagram in Fig.6.9

shows isotropic coacervates up to a critical salt concentration c_{NaCl}= 800mM. The comparison with the previous case asserts that on one hand protamine-single strand DNA interaction increases the critical salt concentration PLL-single strand interaction, and on the other hand it reduces their thermal stability in the narrow salt concentration range $450 < c_{NaCl} < 600$mM.

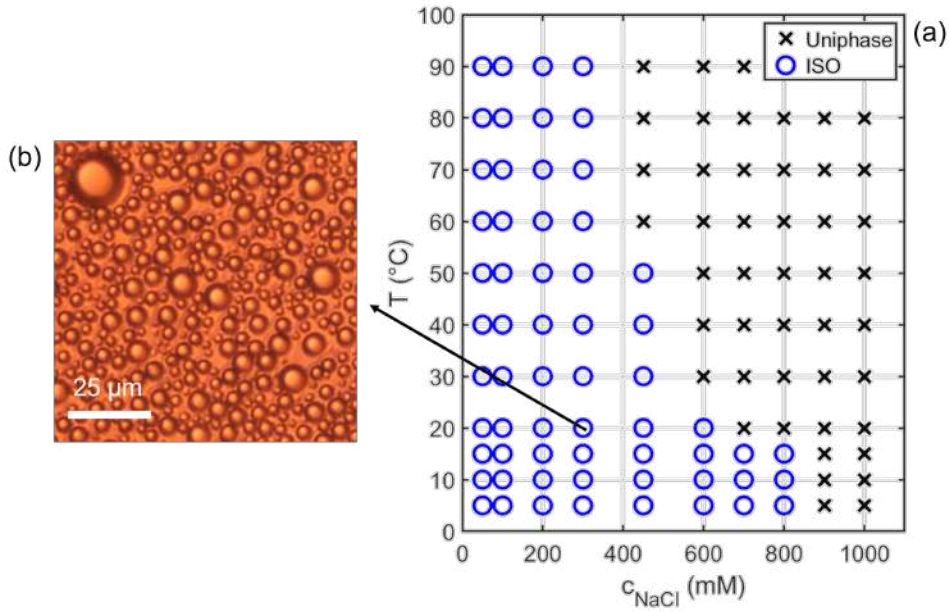

Figure 6.9: Panel(a). c_{NaCl}-T phase diagram for aqueous solution of 12A and protamine, obtained upon cooling from 90°C and PTOM image corresponding to salt concentration and T indicated by arrow. Panel(b). Example of isotropic coacervate.

6.4 Conclusions

In this preliminary work I explored the interaction of several combinations of polycations and nucleic acids. In particular, the interactions of the four polycations with DD duplexes lead to different phase diagrams. Indeed, the smallest polycations with the least charge, spermine and cobalthexamine, lead to the formation of tiny isotropic coacervates (few microns) in a wide range of salt concentrations. Differently, the big and highly charged PLL and protamine give rise to larger isotropic coacervates (up to $\approx$ 20 microns) and LLCPS.

Such polycations have been also investigated with a single strand DNA and 12N DNA oligomers. Among all polycations, protamine unveiled to induce LLPS in a wider range of salt concentration with all DNA oligomers. Differently, the PLL is more effective in inducing LLCPS with also higher thermal stability, as highlighted by phase diagrams of DD-PLL and 12N-PLL.

LLCPS has been also observed for the aqueous solution of DD-protamine, in a wider range of c_{NaCl} compared to PLL, and for DD-spermine only at very low temperature and salt concentration.
The onset of LC in PLL-12 coacervates nicely summarises some of results of this PhD thesis. Indeed, in Chapter3 it has been shown that in random solution at low T, most duplexes have less than two pairing errors with high percentage of defectless duplexes ($\approx 15\%$). If duplexes were highly defected, it's likely that no LC ordering could occur at any concentration, as for single strand nucleic acids. How these hybridized duplexes interact each other to sustain LC order and how PLL inside coacervates modifies such interactions to give rise to higher ordered columnar LC phase is a really fascinating topic that will be investigated in future works.
LLCPS in aqueous solution of 12N-PLL involves exclusively biological molecules that could have been found in prebiotic context where random nucleic acids were inevitably present. Thus, such outstanding result further enhances the hypothesis of liquid crystal ordering as favourable route for the onset of long nucleic acids[16, 17] capable to support biological activity.

Bibliography

[1] Yongdae Shin and Clifford P Brangwynne. Liquid phase condensation in cell physiology and disease. *Science*, 357(6357), 2017.

[2] Karina K Nakashima, Mahesh A Vibhute, and Evan Spruijt. Biomolecular chemistry in liquid phase separated compartments. *Frontiers in molecular biosciences*, 6:21, 2019.

[3] HG Bungenberg de Jong. Complex colloid systems. *Colloid science*, 2:335–432, 1949.

[4] Dimitrios Priftis, Nicolas Laugel, and Matthew Tirrell. Thermodynamic characterization of polypeptide complex coacervation. *Langmuir*, 28(45):15947–15957, 2012.

[5] Erica A Frankel, Philip C Bevilacqua, and Christine D Keating. Polyamine/nucleotide coacervates provide strong compartmentalization of mg2+, nucleotides, and rna. *Langmuir*, 32(8):2041–2049, 2016.

[6] Tommaso P Fraccia and Tony Z Jia. Liquid crystal coacervates composed of short double-stranded dna and cationic peptides. *ACS nano*, 14(11):15071–15082, 2020.

[7] Sarah L Perry, Yue Li, Dimitrios Priftis, Lorraine Leon, and Matthew Tirrell. The effect of salt on the complex coacervation of vinyl polyelectrolytes. *Polymers*, 6(6):1756–1772, 2014.

[8] Jeffrey R Vieregg, Michael Lueckheide, Amanda B Marciel, Lorraine Leon, Alex J Bologna, Josean Reyes Rivera, and Matthew V Tirrell. Oligonucleotide–peptide complexes: phase control by hybridization. *Journal of the American Chemical Society*, 140(5):1632–1638, 2018.

[9] Anisha Shakya and John T King. Dna local-flexibility-dependent assembly of phase-separated liquid droplets. *Biophysical journal*, 115(10):1840–1847, 2018.

[10] Arnab Mukherjee, Ambroise de Izarra, Jeril Degrouard, Enrick Olive, Prabal K Maiti, Yun Hee Jang, and Yves Lansac. Protamine-controlled reversible dna packaging: A molecular glue. *ACS nano*, 15(8):13094–13104, 2021.

[11] Frederic Poineau, Edward Mausolf, William Kerlin, and Kenneth Czerwinski. Hexaammine-cobalt (iii) pertechnetate: preparation, structure and solubility. *Journal of Radioanalytical and Nuclear Chemistry*, 311(1):775–778, 2017.

[12] Marco Todisco, Gregory P Smith, and Tommaso P Fraccia. Liquid crystal ordering of dna dickerson dodecamer duplexes with different 5'-phosphate terminations. *Molecular Crystals and Liquid Crystals*, 683(1):69–80, 2019.

[13] Nupack. http://www.nupack.org/partition/new.

[14] Yves Lansac, Jeril Degrouard, Madalena Renouard, Adriana C Toma, Françoise Livolant, and Eric Raspaud. A route to self-assemble suspended dna nano-complexes. *Scientific reports*, 6(1):1–12, 2016.

[15] King JT Shakya A, Andresen Kurt, S Lamb Jessica, W Kwok Lisa, and Pollack Lois. Dna local-flexibility-dependent assembly of phase-separated liquid droplets.e. *Biophys J*, 2018.

[16] Marco Todisco, Tommaso P Fraccia, Greg P Smith, Andrea Corno, Lucas Bethge, Sven Klussmann, Elvezia M Paraboschi, Rosanna Asselta, Diego Colombo, Giuliano Zanchetta, et al. Nonenzymatic polymerization into long linear rna templated by liquid crystal self-assembly. *ACS nano*, 12(10):9750–9762, 2018.

[17] Tommaso P Fraccia, Gregory P Smith, Giuliano Zanchetta, Elvezia Paraboschi, Youngwooo Yi, David M Walba, Giorgio Dieci, Noel A Clark, and Tommaso Bellini. Abiotic ligation of dna oligomers templated by their liquid crystal ordering. *Nature communications*, 6(1):1–8, 2015.

Acknowledgments

Ringrazio Tommaso B. per avermi dato la possibilità di lavorare nel suo gruppo di ricerca e per la sua instancabile curiosità e passione per la scienza.
Ringrazio Tommaso F., in particolare per le due bellissime e intense settimane di lavoro a Parigi.
Ringrazio Giovanni per la sua infinita pazienza nel debuggare lo scattering e per le sue scary doti da hacker.
Ringrazio Stefano, per la sua positività e allegria e per questi anni di lavoro insieme di cui sono profondamente grato.
Ringrazio il buon Scotti per la compagnia, le pause caffè e il supporto vicendevole nei momenti di difficoltà sperimentali.
Ringrazio tutto il gruppo dei complessi per la compagnia in questi anni. In particolare Giuliano e il suo indimenticabile "ad avere tempo e voglia infinita...", Fabio, Roberto, Marco B., Eddy, Savo, Manuèl, Marco T., Federico, Francesco, Laura, Smile, Luka, Luchino e Thomas!
Ringrazio i miei genitori e mio fratello Ale, in particolare per il vostro sostegno nel primo anno di dottorato.
Ringrazio Luciana e Nicola e i nonni MariaRosa e Pepe, il vostro aiuto è stato prezioso per potere scrivere questa tesi nelle vacanze di Natale.
Grazie ai miei amici, Luigi, Betta, Elena, Franci, Davide, Tino e Marghe, Martina, Giacomo, Ale, Mambro e Fede, Cesare, per avermi continuato a indicare dove guardare.
E infine un ringraziamento speciale alla mia famiglia: a mia moglie Nicol, per il suo continuo e amorevole sostegno e alla nostra piccola Teresa, i suoi sorrisi quotidiani donano gioia al mondo.