UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Fisica, Astrofisica e Fisica Applicata

Dipartimento di Biotecnologie e Medicina Traslazionale

Corso di Dottorato in Fisica, Astrofisica e Fisica Applicata

Ciclo XXVI

Exploring the role of Liquid Crystal ordering of DNA oligomers in the prebiotic synthesis of nucleic acids

Settore Scientifico Disciplinare FIS/07

Supervisore: Professor Tommaso BELLINI

Coordinatore: Professor Marco BERSANELLI

Tesi di Dottorato di:

Tommaso Pietro FRACCIA

Anno Accademico 2013-2014

Commitee of the final examination:

External Referee:
Prof. Dieter BRAUN

External Member:
Prof. Paolo MARIANI

Internal Member:
Dott. Roberto CERBINO

Final examination:

Date Feb 13, 2013

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

To Giulia, Mauro
and to my family

There are more things in heaven and earth, Horatio,
Than are dreamt of in your philosophy

W. Shakespeare, Hamlet (1.5.167-8)

Cover illustration:

Spherulites growth from DNA 4mer

MIUR subjects:

FIS/07

PACS:

87.14.-g

Contents

Abstract

The work in this thesis has been devoted to study the self-assembly of DNA oligomers in light of their possible relevance in the context of the formation of longer chains of nucleic acids in the prebiotic world. This conjecture is based on previous evidence of hierarchical self-assembly of short oligonucleotides in solution that provides mechanisms of self-selection. In such self-assembled structures, the 3′ and 5′ terminals are held in close contact, a condition that could act as a spontaneous template for the elongation of the chains in conditions favoring their chemical ligation.

The work here described was aimed at testing this notion by various investigations targeting what appeared to be the most critical issues in this context. Accordingly, I acted in three main directions.

First I investigated if liquid crystal (LC) ordering, and hence ligation templating, emerges even in solutions of oligomers having sequences chosen at random, and I determined a new extended phase diagram which includes the degree of randomness. Second, I determined what is the minimum oligomer length which enables the formation of LC phases and I found that LC phases can be found in solution of oligomers as short as 4-bases, and even with randomly chosen sequences. This last result was actually quite suprising and indicates the existence of a new regime for the self-assembly of ultra-short DNA chains. Third, I explored the influence of LC ordering of short DNA oligomers on non-enzymatic ligation reaction favored by the presence of a water-soluble condensing agent. I found a good yield for the polymerization of DNA oligomers both in isotropic than in LC phases, and polymerized chains up to 12 times longer than the initial length.

I believe the work described in this thesis strengthens the notion that self-assembly of nucleic acids could indeed have been the key factor promoting the formation of long chains. These results shad a new light on the most obscure among the processes that enabled the emergence of life on the early Earth.

The work described in this thesis is partially published in:

T. Bellini, G. Zanchetta, T. P. Fraccia, R. Cerbino, E. Tsai, G. P. Smith, M. J. Moran, D. M. Walba and N. A. Clark, *Liquid crystal self-assembly of random-sequence DNA oligomers*, Proc. Natl. Acad. Sci. USA **109** (2012) 1110

Introduction

It has been recently shown that short DNA oligomers (down to 6 bases long) associated in double helices may order into Liquid Crystal (LC) phases despite their nearly globular shape [1]. In these systems the formation of LC is mediated by the end-to-end aggregation of DNA duplexes into columns of chemically distinct but physically continuous duplexes. When the axial ratio of such self-assembled columns fulfills Onsager criteria for LC ordering, they develop either nematic or columnar LC phases.

Further investigations have revealed that the process of LC formation is selective, such that in mixture of complementary and non-complementary oligomers, the complementary oligomers tend to be incorporated into the LC domains, and segregated from the surrounding isotropic phase of non-complementary strands. LC phases are found even in systems in which double stranded DNA is mixed with poly-(ethylene glycol) (PEG) chains. In both these systems the formation of LC domains is associated to a condensation of stacked DNA duplexes. [2]

The origin of the research conducted in the frame of this thesis was the intriguing hypothesis that this combination of self-assembly steps leading the formation of LCs, unique of DNA molecules, could have been relevant for the very origin of the nucleic acids themselves.

Indeed the chain of events that led to the emergence of life from the prebiotic Earth is still lacking of a convincing scientific proof. The Miller experiment and recent repetitions of similar experiments showed that several biologically relevant organic molecules (such as amino acids and nucleobases) could be produced by a variety of conditions, including the famous Miller's electric discharge on reducing gas mixture [3, 4]. However, how life could have emerged from such mixture of small carbon-based molecules is still unknown. Nowadays one of the most prominent theory is the RNA world proposed by G. F. Joyce, according to which the first element of life was a combination of RNA strands that were capable of self catalyze its own reproduction [5]. This notion is based on various clues, the most important being the observation that specific sequences of RNA fold and become capable of enzymatic-type activity. These chemically active catalytic RNA structures are called ribozymes. Thus, it is possible to imagine a situation where a specific pool of such ribozymes could act on simpler carbon-based molecules and lead to the

synthesis of more ribozymes. In order to become catalytic, RNA polymer chains must be at least 50 monomer units in length. Thus, in the frame of the RNA world hypothesis, the question about the origin of life transforms into: "How could such RNA strand have been formed in completely non-biotic way?". Iterations of random chemical reactions are nowadays considered too improbable to convincingly be able to have produced such complex molecules.

As the Nobel laureate C. De Duve recognizes [6], the need of a "guiding hand" seems inescapable to address the reactions from single nucleotides towards the polymerization of nucleic acids. Could such "guiding hand" have been provided by the self-assembly capacities of nucleic acids? According to this vision, nucleic acids could have had peculiar characteristics that catalyzed their own formation and promoted their supremacy against other possible molecular forms, in order to survive in a selective environment and to become the carriers of the genetic code of life.

The idea that hierarchical self-assembly of short DNA and LC ordering could represent a crucial property has always intrigued my supervisor, prof. Tommaso Bellini, and prof. Noel Clark (University of Colorado at Boulder). Their enthusiasm soon infected me and other people in their research groups. This thesis is the result of our research focused on considering and testing DNA self-assembly as a possible route for the prebiotic synthesis of nucleic acids. Indeed, the molecular ordering within LC domains, in which the terminals of distinct oligomers are held in close proximity by packing and stacking forces, could have acted as a template for the formation of extended chains, if the chemical conditions of the solution favored the ligation of such terminals. Moreover the strong link between liquid crystallization and segregation, even in crowded environment, acts as a powerful selection mechanism that represent a new peculiarity of nucleic acids.

The work described in this thesis is aimed at testing the potential of this hypothesis. To do this, I devoted my activity along three main directions to tackle what seemed to be the most difficult aspects of this concept. In other words, I tried to push the limits of the observed self-assembly of nucleic acids toward conditions that appear less improbable to achieve.

First, I investigated the behavior of random sequences DNA, in the notion that prebiotic nucleic acids could not have been randomly produced with definite sequences.

Second, I investigated the behavior of shorter sequences, since in a context or random chemistry these appear as more probable.

Third, I examined what conditions could have promoted the transformation of physical proximity of chain terminals into their chemical bonding.

Part of my thesis work was devoted to study the behavior of solutions of DNA oligomers in which the base sequence are chosen at random, or with varying degrees of randomness. I showed that even in such a crowded environments, like solutions of completely random sequences, corresponding to a remarkably large number of different molecules, e.g., $\sim 10^{12}$ for random 20mers, complementary still emerges and, for a narrow range of oligomer lengths ($16 \leq N < 30$), leads to structured self assembly in the

form of organization of the single strands into LC phases. This ordering follows from a subtle hierarchical self-assembly sequence in which the kinetic arrest of oligomer association into long-lived partially paired double helices is followed by reversible association of these pairs into linear aggregates that condense into LC domains.

The presence of a lower limit to the DNA length for LC phases ($N = 16$ for random and $N = 6$ for definite sequences) formation seemed to invalidate the prebiotic relevance of these systems. However recently I found LC ordering of 4-base-long DNA oligomers, i.e. two bases less than previously observed. LC ordering of 4mer with definite sequences has been widely characterized in this thesis work and concentration-temperature phase diagram has been obtained. These findings revived the investigation on the minimum oligomer length which enables the formation of LC phases, the study of oligomers shorter than 4 bases is still a work in progress.

Moreover I have been surprised by finding re-emergence of LC ordering also with very short random sequences, i.e. random 4mers. In this case, unlike the previous model for self assembly of random oligomes, due to the shortness of the components, the complete oligomer association is kinetically accessible. Nevertheless a slightly different mechanism drives the ordering in this regime of lengths in which there is no significant difference between two-strands pairing, that leads the double helix formation, and multiple-strands interaction that favors a cooperative linear chain formation.

At the same time, I have tested the influence of LC ordering of short DNA oligomers on non-enzymatic ligation reaction. Specifically I have investigated the ligation of DNA Dickerson dodecamer with a 3'-phosphate termination in the presence of EDC, a water-soluble carbodiimide as condensing agent. Despite having adopted the choice of not introducing modified molecular terminals, such as the highly reactive amino-termination [7], I found a good yield for the polymerization of DNA oligomers both in isotropic than in LC phases. Quite remarkable is the fact that the LC condensation in mixture of DNA oligomers with PEG causes a sharp increase in the degree of polymerization up to lengths of 144 bases (12 times the initial length). This shows how ordering and compartmentation of DNA double strands has a crucial role in catalyzing non-enzymatic ligation reactions.

The chapters in the thesis are organized as follows:

Chapter 1 contains an introduction to the structure and the main properties of DNA and brief description of the tight link between DNA and Liquid Crystal mesophases.

Chapter 2 gives an introduction to the origin of life problem and it reviews the main scientific advancements in the last years.

Chapter 3 describes materials and experimental techniques adopted in this research.

Chapter 4 reports about liquid crystal self-assembly of random-sequence DNA oligomers.

Chapter 5 discusses the LC ordering of 4 bases long DNA oligomers and its phase diagram.

Chapter 6 shows our results on ligation of short DNA oligomers in LC phase.

CHAPTER 1

DNA and Liquid Crystals

In 1953 Watzon and Crick discovered the double helix structure of the DNA with the fundamental contribution of X-rays diffraction pattern. It has been recognized that the famous "photo 51" of Rosalind Franklin was taken with DNA in a nematic liquid crystal state, a lucky strike that, hiding the diffraction patterns of the crystalline lattice periodicity, allowed the isolation of the structural information coming exclusively from the the double helix itself. The LC ordering was due to the sufficiently large length and rigidity of the DNA double helices. Since that moment the story of DNA and of liquid crystals had several intriguing meeting points which even suggest a deeper link between these two apparently far worlds.

In this Chapter we introduce the structure and the main properties of DNA and we give a brief description of the DNA Liquid Crystal mesophases.

1.1 DNA structure and fundamental interactions

DNA, deoxyribonucleic acid, the most celebrated molecule of life, is a polymer that usually associates in pairs to form the famous double helix, in which the two strands are wound around each other. DNA is a polynucleotide that is a chain of repeated basic units (nucleotides) comprised of a sugar deoxyribose, a nitrogen base (nucleobase), and a phosphate group. There are four types of nucleotides in DNA, which only differ in the nucleobase: adenine (A), guanine (G), cytosine (C), and thymine (T). Each polynucleotide is hence characterized by a specific nucleobase sequence, conventionally written in the direction from the 5'-end to the 3'-end of the sugar phosphate backbone. The double helix is shaped as a spiral "staircase" whose steps are the individual nucleotides. The right-handedness of the helix is a consequence of the chirality conformation of the sugar moiety. The sugar-phosphate backbones of the two DNA strands are antiparallel with the 30 terminal of one strand connected to the 5' terminal of the other (see Fig. 1.1). In the most common structure, the B-form, the bases remain on average perpendicular to the molecule axis. Helix periodicity is ~ 10 base pairs, equivalent to $\sim 3.4nm$, while the diameter of the bare DNA is $\sim 2nm$ [8].

The sequence of nucleobases is what gives DNA its capacity of carrying the genetic information and this is why it is crucial that the nucleotides form covalent bond chains. By contrast, the two strands are held together by weak interactions to allow for the zipping and unzipping of the double helix, the critical step involved in the "reading" of the

sequence during transcription and replication of DNA in the cell. Indeed, the biological role of DNA is rooted in the physical properties of its molecular structure. The solubility, flexibility, and mutual interactions of DNA strands are key to its capacity to encode and transfer biological information. The very shape of the double helix is the result of a subtle balance of molecular constraints and interaction forces, all playing a role in the biological processes. Double helices are held together by a combination of two main intermolecular forces: the so-called base pairing and base stacking forces. These forces control the selectivity of the binding process, by which the binding energy depends on the degree of complementarity of the nucleobase sequences along the two polymers.

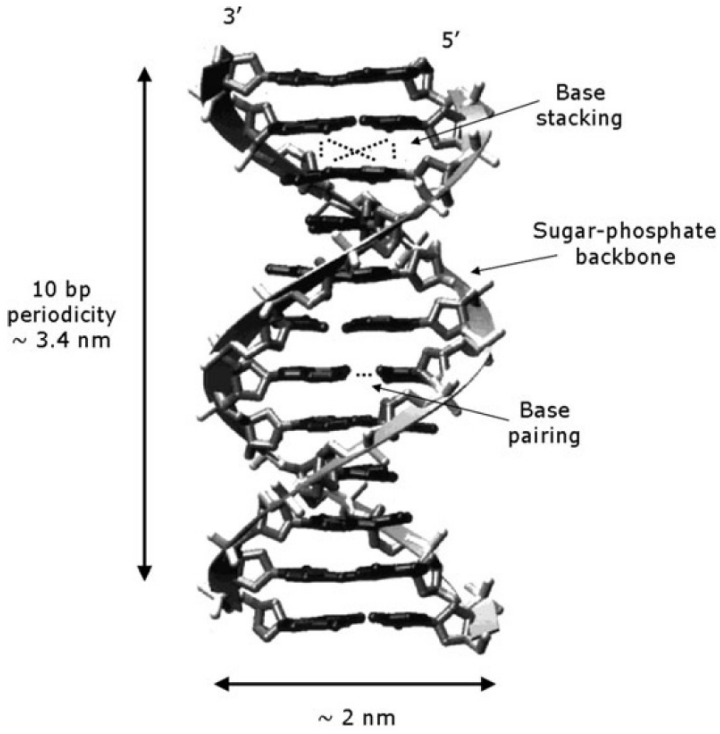


Figure 1.1: The double helix structure of DNA in its B-form, with its relevant dimensions and the main interactions. From [9]

When instead assemblies of helices are taken into account, it is well known that for many aspects DNA duplexes in solution can be treated as a charged anisotropic particle [10]. Accordingly, steric, electrostatic, and Van der Waals interactions, together with the mechanical properties of the helix (bending and torsional rigidity), play a major role in the formation of DNA mesophases. In addition, all these different kinds of interactions combine in a subtle and still poorly understood way to generate other forces relevant for the case of DNA.

1.2 Thermodynamic Stability of the Double Helix

Base pairing originates from the formation of hydrogen (H) bonds between pairs of nucleotides. The strength of the interaction is maximum when the sequences match according to the Watson–Crick (WC) complementary couples, i.e., A-T (two H bonds) and C-G (three H bonds). The energy associated with the formation of the double helix ("hybridization") depends on the length of the nucleic acid polymer and on the quality of the WC matching, vanishing for pairs of sequences with poor complementarity. Despite the fact that the conceptual description of the pairing of complementary bases is simple, the quantitative evaluation of the energy involved is complicated because, when unpaired, the nucleobases form H-bonds with water.

Base stacking forces, instead, act to pull the aromatic hydrocarbon plates of the nucleobases into contact with each other. It is believed that this is due to the hydrophobic nature of the base surfaces whose exposure to water tends to be minimized. Stacking is much less selective than pairing and its strength mainly depends on the overlap between the aromatic plates. Nevertheless stacking forces are the main sources of stability of the double helix, while pairing provides the necessary specificity of the bonds. When combined with the constraint of the sugar-phosphate backbone, the stacking forces result in the twist of a base pair relative to the neighboring ones of $\sim 36^\circ$.

The possibility of predicting a priori the interaction strength between two or more nucleotides sequences, together with the availability of the highly developed chemical machinery to synthesize sequences on design, makes nucleic acids very appealing to design molecular and supramolecular structures mutually interacting in a controlled way. The number of experimental systems constructed in this way is rapidly growing and this is at the same time the cause and the consequence of the large amount of quantitative thermodynamic studies on the thermal stability of double helices. In the past, experimental results obtained at various concentrations have been analyzed in the attempt to extract the amplitude of pairing and stacking forces. However, since pairing and stacking always act together in determining the DNA structure, it is hard to separate their contributions. Hence, thermodynamic studies of DNA are typically focused on determining the combined value of stacking and pairing in the overall Gibbs free energy G. Here the interest is in the free energy difference $\Delta G = G_D G - U$, where G_D and G_U refer to duplexes and unbound strands, respectively. It is generally found that the formation of a DNA duplex involves variations of both enthalpy (ΔH) and entropy (ΔS), where $\Delta G = \Delta H - T\Delta S$. The large body of observations has enabled the formulation of simplified approximate strategies to express ΔG given the two sequences involved.

The most commonly adopted approach, the "nearest neighbor" (NN) model [11, 12], is based on a very simple idea: since pairing and stacking always act together, the minimum entity which contains both effects is a quadruplet formed by two consecutive nucleotides on one strand and the corresponding nucleotides on the other strand. This model and its use in the present work are widely discussed in Section 3.6.

1.3 DNA melting temperature

Despite the lack of precision in determining the various contributions yielding the actual free energy of the DNA helices, the overall values of ΔG, ΔH, and ΔS for each given sequence can be experimentally determined with good accuracy through melting experiments. As T is increased, the weak non-covalent bonds break and the double helices become unstable and denaturated. The melting temperature T_m is defined as the temperature where, for half of the double helices, the two strands have unbound and split, i.e., where $\Delta G = 0$. Such a transition can be monitored by optical (absorbance, fluorescence) or calorimetric methods. An example is reported in Fig. 1.2, where the UV absorbance and the corresponding fraction of single strands are plotted as a function of the temperature T. T_m is determined as the middle point of single strands to double helix transition. Experiments indicate that for a fixed ionic strength T_m is very sensitive to the number N of base nucleotides for short oligomers, while for long double helices it reaches a saturated, yet rather high, value.

1.4 DNA liquid crystals

Liquid crystals (LC) are fluid phases characterized by a partial orientational and positional ordering, and are thus normally classified as mesophases, in between the isotropic liquid state and the crystalline packing. Initially viewed as an odd form of collective ordering of a very limited set of organic molecules, LCs are nowadays recognized as a common form of self-organization of soft and biological matter. This trend has also been followed by the study of LC ordering of DNA. DNA LCs were first observed in vitro with long double strands; later they were recognized as the in vivo packing mechanism of some organisms, and quite recently they started being considered as a common ordering for DNA oligomers.

1.4.1 Long DNA

Early studies carried out between the 1950s and the 1960s identified mesophase behavior of long DNA chains of biological origin in vitro [14, 15]. Shortly thereafter Bouligand started a systematic analysis of the textures of various biological specimens obtained through optical and electronic microscopy, in particular of the chromosomes of Dinoflagellates (primitive unicellular algae) finding evidence of LC arrangement of highly concentrated DNA in vivo [16]. LC order of DNA was also found in other chromosomes, bacteria, viruses, and sperm heads, where its concentration reaches extremely high values, up to $800mg/ml$, suggesting a correlation between LC packing and biological activity [17, 18], in particular with respect to protection from external stress or damage [19].

DNA LC phases have been extensively studied by using more easily controllable experimental systems, namely DNA double strands obtained through enzymatic cut of nucleosomal DNA, providing a nearly monodispersed ensemble of $N \sim 150$ semi-rigid rods, or through sonication, which yields average lengths N between $10^2 bp$ and $10^4 bp$ (contour length L between $30nm$ and $3mm$) with relatively narrow distributions.

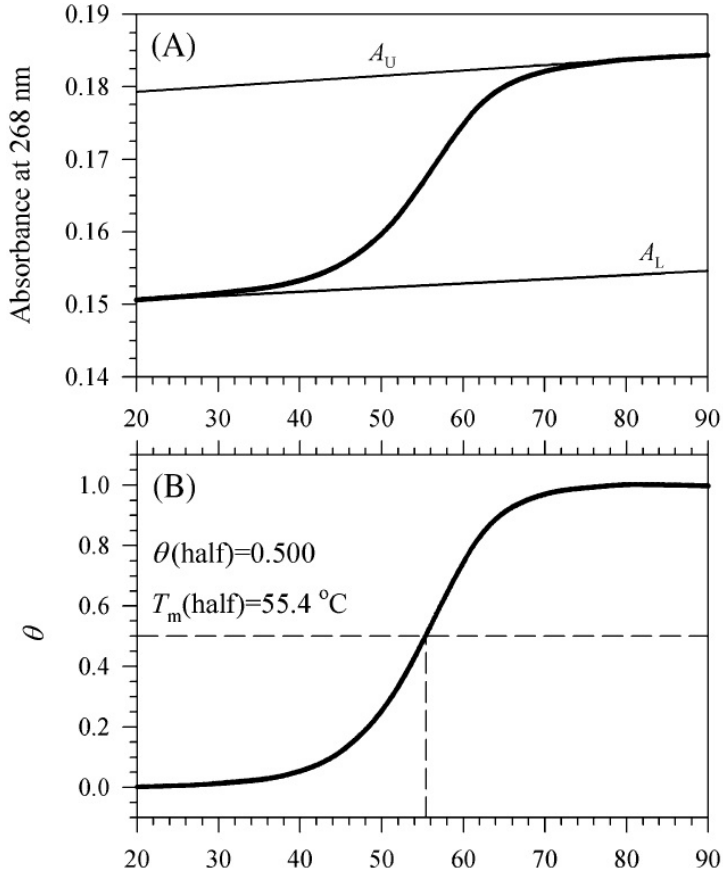


Figure 1.2: Thermal denaturation transition of a DNA helix. (a) UV absorbance increases with temperature, following the unstacking of bases, following a sigmoidal shape. A_D and A_U are lower and upper baselines, also slightly dependent on temperature. (b) Fraction of single strands θ extracted from data in (a), which defines the melting temperature corresponding to $\theta = 0.5$. Adapted from [13].

Two main mesophases were identified:

1. A cholesteric, or chiral nematic (N*) phase. This is a positionally disordered fluid in which the constituent molecules align on average their axes along a common direction called the nematic director. Being the DNA helices chiral, the orientational order develops an additional macro-helical superstructure with the twist axis perpendicular to the local director. The phase thus consists of local nematic "layers" continuously twisted with respect to each other, with periodicity $p/2$ (where p is the cholesteric pitch; see Fig. 1.3) [20, 21]. For $150-bp$ helices, the N* phase appears at a concentration around $150mg/ml$ in 100 mM monovalent salt conditions. This LC phase is easily observed in polarized optical microscopy. Since the N* pitch extends to tens of micrometers (that is, across more than 10^4 molecules), the optical anisotropy of DNA bases leads to characteristic textures such as the droplet

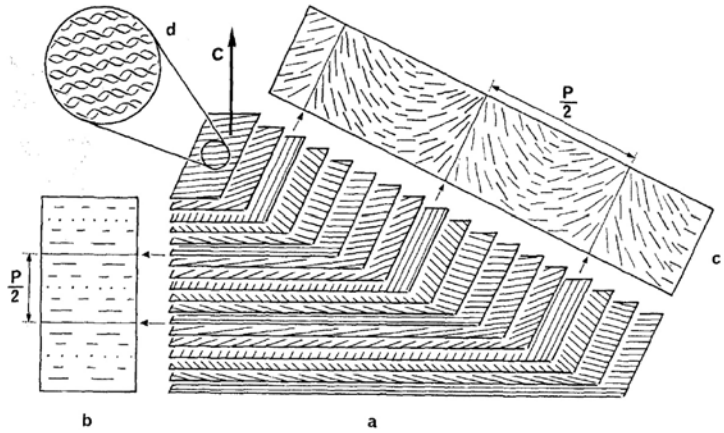


Figure 1.3: Schematic representation of the structure of the chiral nematic phase of DNA, showing continuously twisting nematic layers, giving rise to a $p/2$ periodicity easily observable in the side view on the left (from [20]).

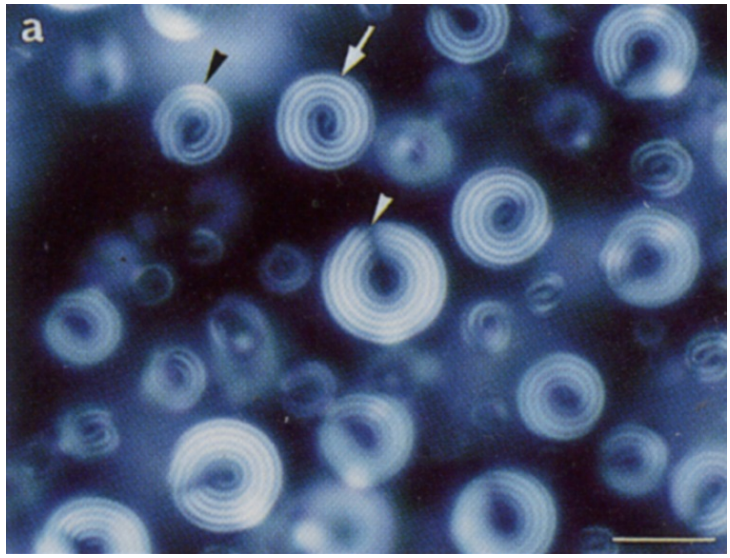


Figure 1.4: N* droplets observed in polarized microscopy. The bright and dark stripes correspond to $p/2$ (from [20]).

in Fig. 1.4, where the stripes correspond to the macrohelical periodicity of the N* ordering.

2. A columnar (COL) phase, in which parallel DNA helices align on a 2D lattice (see Fig. 1.5) but remain free to slide relative to each other in the orthogonal direction. The continuous bending of the columns gives rise to the so-called developable domains, shown in Fig. 5.5. The nature of the phase and its hexagonal packing symmetry were demonstrated by a combination of polarized microscopy, X-ray

diffraction, and freeze fracture electron microscopy experiments [22, 23]. The COL phase is observed for concentrations higher than $400mg/ml$.

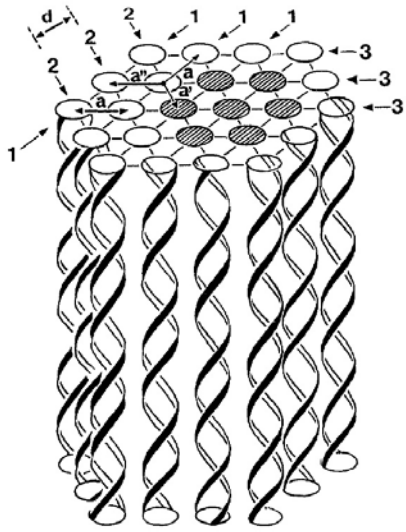


Figure 1.5: Sketch of the structure of the hexagonal columnar phase of DNA, showing parallel molecules hexagonally packed in the plane perpendicular to their axis. "a" and "d" are the lattice parameters. (from [20]).

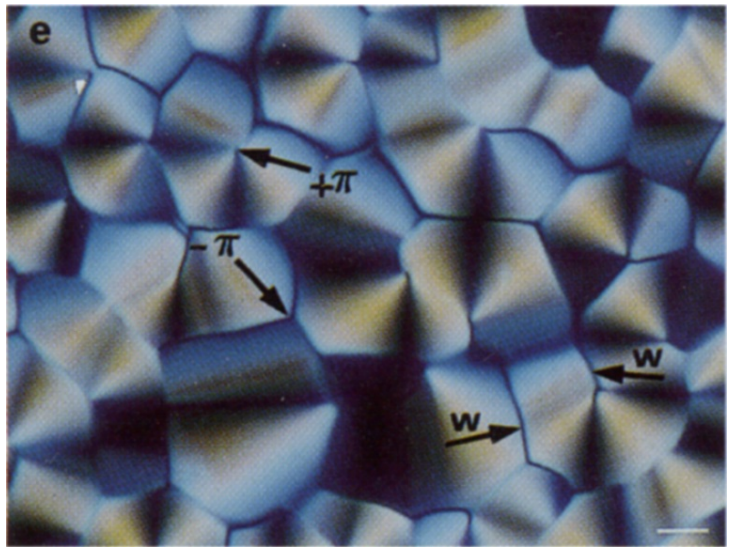


Figure 1.6: COL developable domains observed in polarized microscopy. "w" indicates defect walls between differently oriented domains, while "$\pm\pi$" stands for point defect around which DNA molecules continuously bend (from [20]).

Experimental evidence was reported for the existence of various additional phases: a pre-cholesteric order in the form of a network of double-twisted cylinders, analogous to the thermotropic blue phases [20], a hexatic phase that replaces the hexagonal columnar in very long DNA fragments [24], and a structure with orthorhombic symmetry appearing in the transition to crystalline order [20].

The appearance of a DNA nematic phase follows from it basic rod-like shape and its rigidity when in helical conformation. While single-stranded DNA is extremely flexible, with a polymer persistence length as short as $2nm$ [25], when paired into the double helix it becomes a rather stiff object, with persistence length around $50nm$ (that is, ~ 150 base pairs) [26]. According to the classic argument by Onsager [27] based on excluded volume interactions, elongated semi-flexible repulsive rods at sufficiently high concentration are expected to develop orientational order. The nematic order becomes favored because the loss of orientational entropy is more than compensated by the increased positional entropy. Despite its polydispersity, flexibility, and electrostatic repulsion, DNA proved to be a good test-bench for the Onsager prediction [27]: once the effective diameter D is properly rescaled to include repulsive interactions (in turn depending on the ionic strength), the critical volume fraction for nematic ordering is found to match Onsager's prediction, $\varphi_{IN} \approx 4D/L$ [28], where L is the length of the double helix (see Fig. 1.7).

In a hard-rod system, at sufficiently high volume fraction a transition is usually expected from the nematic to the smectic A phase [29], a lamellar phase with layers perpendicular to the nematic director. However, as elegantly demonstrated by Livolant [22], in DNA the smectic phase is replaced by columnar ordering; this behavior can easily be explained on the basis of strand flexibility [30] or length polydispersity [31], both favoring the COL phase over smectic.

High concentration, however, is not the only means to obtain LC phases of DNA in aqueous solutions. DNA has been found to collapse upon adding in the solutions various condensing agents, which introduce effective attractive interhelical interactions. This is what happens with alcohols and other solvents, which reduce DNA solubility [32], or with multivalent cations like spermidine, spermine, and cobalt hexamine, which are thought to establish correlated counterion fluctuations with a resulting net attraction between DNA molecules [33]. The most common structure of the resulting aggregates is that of a toroid of hexagonally packed DNA, similar to that observed within virus capsids [34], but the addition of non-adsorbing polymers reproduces the formation of the same mesophases described above for concentrated solutions [10]. Similar behavior is observed for multivalent ions in the case of $150 - bp$ strands [33].

1.4.2 Short DNA

According to the Onsager theory and to computer simulations of the behavior of hard sphero-cylinders [29], in the absence of additional interactions no LC ordering is predicted for rods with $L/D < 4$, and therefore DNA double helices with a number of base pairs $N < 24$ would lack the anisotropy to display mesophase behavior at any concentration (Fig. 1.7).

However, rather unexpectedly, LC ordering has recently been found in concentrated

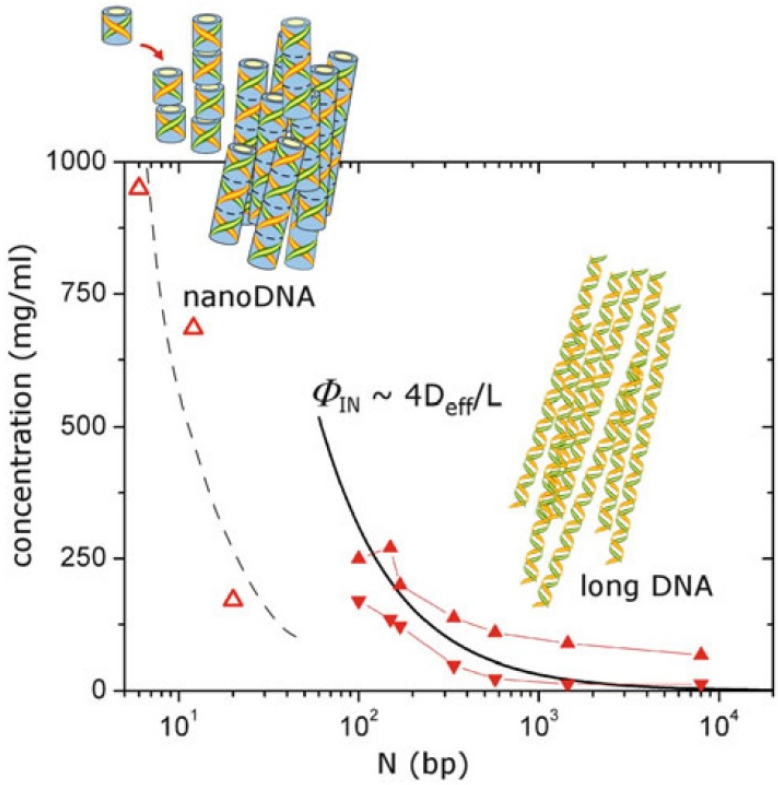


Figure 1.7: Phase behavior of DNA molecules of different lengths. Helices longer than 100 bp (filled triangles) display isotropic-to-nematic transition at concentrations reasonably well described by Onsager theory once the diameter is rescaled for electrostatic repulsion (continuous black line). In contrast, short DNA strands lack the necessary anisotropy to spontaneously align and their critical concentrations (open triangles) imply end-to-end stacking into longer aggregates. The dashed line is a guide to the eye. Long DNA and nanoDNA data are from [28] and [1], respectively. Image from [9].

aqueous solutions of DNA self-complementary sequences with N as low as $6bp$ ("nanoDNA", nDNA), which pair into helices with aspect ratio well below the Onsager limit [1]. The LC phases exhibited by nDNA are the same as those observed in long DNA, N* and hexagonal COL, although they are found at higher concentration, e.g., around $750mg/ml$ for the N* of the self-complementary 12mer. The formation of LC phases of nDNA is a consequence of pairing and stacking interactions.

The first observation of LC in nDNA involved the use of self-complementary sequences. Schematic representation of short DNA self-assembly stages is reported in Fig. 1.8. When these kinds of oligomers pair, they form blunt-ended double helices. The terminal bases of different duplexes tend to stack to avoid exposure to water, leading to an attractive interaction that acts to hold them together when their flat aromatic faces are in contact. These end-to-end interactions induce reversible linear aggregation of the

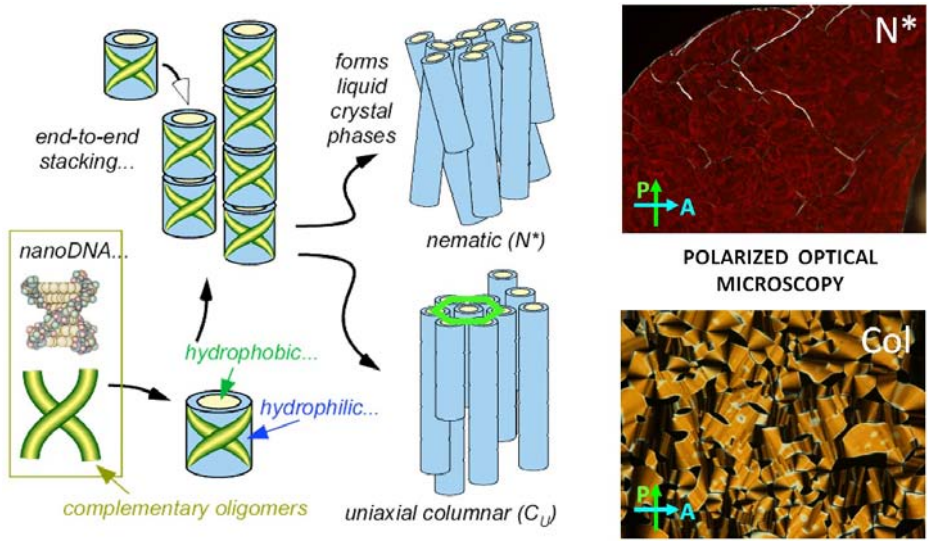


Figure 1.8: Schematic representation of short DNA self-assembly stages. Pairing of complementary strands yield double helices. Helices interact trough end-to-end stacking forces. Successively helices capable of linear aggregation display N* and COL LC phases. Modified from [1].

oligonucleotides into effectively long chains, a process somewhat analogous to the "living polymerization" of surfactants into rod-like micelles (see [35]). When the physically bound, but chemically nicked, chains reach an axial ratio and concentration large enough to enter the Onsager regime, the solution orders into the N* phase, and, upon further increasing the concentration, into the COL phase. The free-energy change associated with the stacking of duplexes, as estimated from the nDNA phase diagram, $\delta G_{BS} \sim 4$–$6 k_B T$, is similar to that expected for base stacking within helices [36]. The freedom of stacked duplexes to rotate around their axis may enable the system to find a better stacking condition (larger stacked surface) than the normal quadruplets within double helices. This fact could indeed explain why the free energy involved in stacked duplexes exceeds that for stacking within helices [37].

The stacking of nucleobases has been found to be involved in several DNA self-assembly processes, promoting the formation of DNA–DNA and DNA–protein complexes in vivo [38, 39], driving end-to-end interactions of DNA double- and triple-helices in semi-dilute solutions [40], or determining the geometry of DNA oligonucleotide crystals [41, 42]. Stacking is also at the origin of the self-assembly processes leading to linear aggregation and LC ordering in solutions of guanosine quartets [43]. We note that in both nDNA and G-quartets the self-assembly is triggered by pairing events: G monomers and DNA single strands have only a weak propensity to stack, but the formation of H-bonds enormously reduces the water solubility of the molecules because of the enlarged hydrophobic surface, and promotes aggregation and ordering.

Since nDNA linear aggregation is critically dependent on stacking interactions, it is easily disrupted by the addition of unpaired bases to one terminal of the selfcomplementary sequence (see T terminations in Fig. 1.9), which is found to suppress LC formation

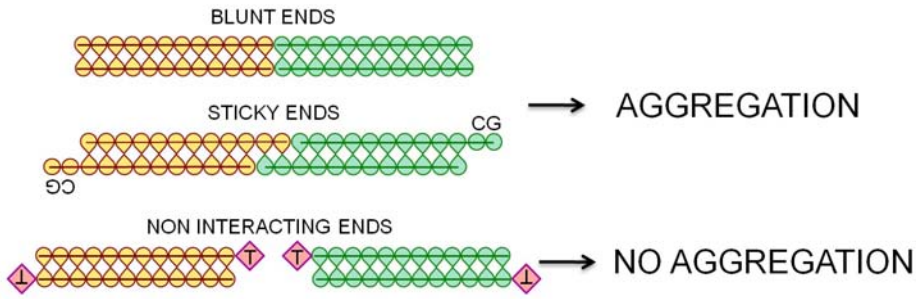


Figure 1.9: Schematic representation of short DNA duplexes which allow aggregation, interacting trough blunt-ends stacking or sticky-ends complementarity, or prevent aggregation, in case of non-complementary dangling ends.

[1]. However, if such single stranded sequences dangling at the end of the nDNA double helix are designed to be complementary to each other ("sticky ends"), the linear aggregation and LC order are restored [44]. An example of sticky end is given by the selfcomplementary sequence CG in Fig. 1.9. Sticky overhangs of different length or sequence and hence different interaction strength can determine very different LC stability. The study of the N* and COL phase boundary thus provides a new direct route to evaluate the strength and the temperature dependence of the stacking and pairing forces [44]. Sticky ends were also recently proposed as an effective driving mechanism to aggregate short DNA strands into longer polymers that are more easily condensed into DNA particles, as a valuable method of oligonucleotide delivery in gene therapy [45].

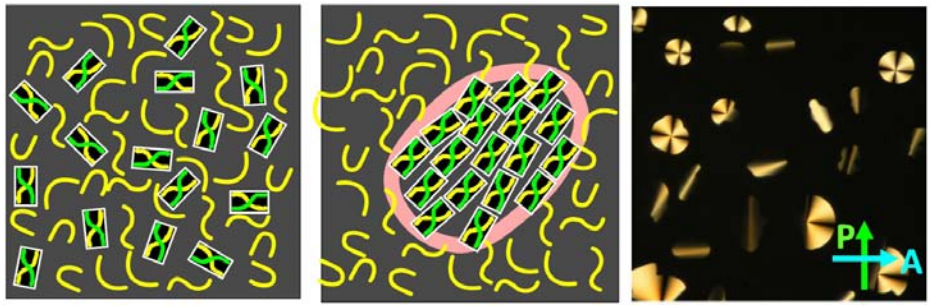


Figure 1.10: Complementary sequences segregate from mixtures of unpaired sequences through the nucleation of LC domains.

The investigation of mixtures of nDNA in which only some of the sequences formed double helices, while others remained in the solution as single strands, led to the recognition of yet another force acting to promote DNA self-assembly. The system was found to undergo phase separation via the nucleation of liquid crystalline domains rich in helices coexisting with an isotropic background rich in single strands, as shown in Fig. 6.13 [2]. The main mechanism driving the phase separation is the difference of flexibility between double- and single-strands. Because of the phase separation, single strands act as an os-

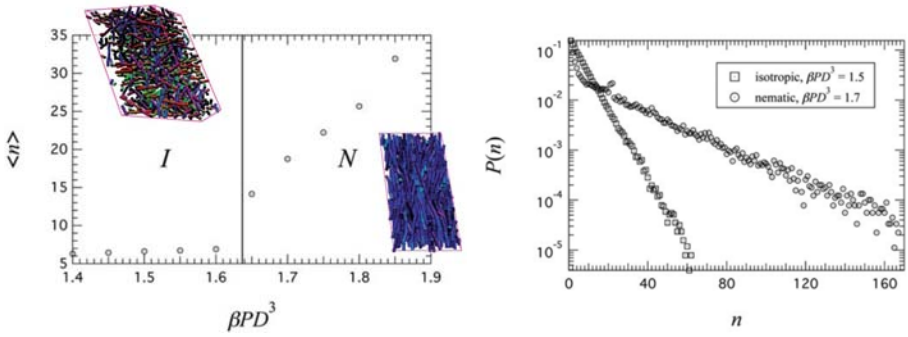
molyte, increasing the local concentration of duplexes above the value required for LC ordering. This behavior is analogous to protein crystallization promoted by dissolving PEG in the protein solution: in that case depletion-type entropic forces develop because of the mismatch between the flexibility of PEG and the rigid structure of proteins in their native state. In the case of DNA the spontaneous partitioning between complementary and unpaired sequences depends on the combination of the entropic depletion-type interactions and the free energy gain from the end-to-end stacking and LC ordering of duplexes [2]. This becomes evident when sequences have a terminal group interfering with the stacking: in this case neither LC formation nor demixing of complementary sequences is observed.

1.5 Linaer aggregation of DNA duplexes

The experimental observations of nDNA LC have also renewed the interest in theoretically modeling systems where reversible linear aggregation and LC ordering are cooperatively correlated [35, 46, 47]. This is also the case for chromonics, flat hydrophobic molecules, mainly dyes, with soluble moieties at their edges [48].

Kuriabova et al. [35] studied a system of sticky cylinders adopting Monte Carlo simulations and analytic theory. Cylinders can interact only by hard-core repulsion and, when they are sufficiently close together, by end-to-end attraction. Isotropic and Nematic phases were investigated by varying three parameters: cylinder aspect ratio L/D, dimensionless pressure $\beta P D^3$ and dimensionless binding energy βE_{bond}. In the isotropic phase the monomers are distributed uniformly both in space and in orientation. Instead in the nematic phase they are uniform in space but orientationally ordered along the direction of nematic director. The measurements of the aggregation number distribution function $P(n)$ and the mean aggregation number $\langle n \rangle$ show the influence of LC ordering the system. As shown in Figure 1.11, $P(n)$ has exponential behaviour in isotropic phase but it broaden to a biexponential in the nematic phase, reflecting the growth of longer aggregates in the nematic phase. The mean aggregation number shows a discontinuity at the I-N transition and increases in the nematic phase (see Fig. 1.11).

Similar results have been obtained by De Michele et al. [46] from numerical simulations on a coarse grained model of DNA duplexes, consisting of bifunctional patchy particles with anisotropic shape. In Figure 1.12 the average chain length (M) vs volume fraction (ϕ) is reported. M grows at increasing volume fraction and it shows a huge jump at isotropic-nematic transition. Moreover larger increase is found for monomers with lower end-to-end binding energies. This behaviour enlighten the effect of orientational ordering of the system on the aggregation quality, that is particularly influential for duplexes with low end-to-end interaction.



(a) (b)

Figure 1.11: Mean aggregation number $\langle n \rangle$ as a function of pressure (a) and aggregation number distributions $P(n)$ (b) obtained with Monte Cralo simulations of sticky cylinders system (sketched in the insets), adapted from [35]. $\langle n \rangle$ increases passing from isotropic to nematic phase with an evident jump at I-N transition. $P(n)$ has exponential behaviour in isotropic phase but it broaden to a biexponential in the nematic phase.

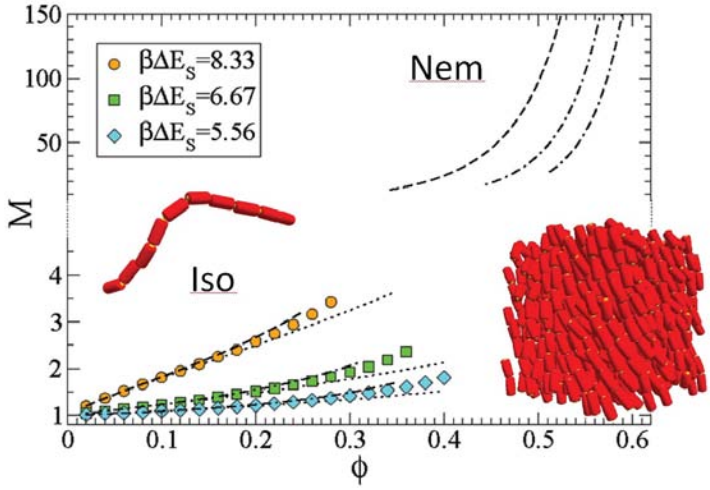


Figure 1.12: Average chain length (M) vs monomers volume fraction (ϕ) for a system of bifunctional patchy particles with anisotropic shape ($L/D = 2$) for three different binding energies $\beta\Delta E_S$. Symbols are Monte Carlo simulation in the isotropic phase. Dashed lines are theoretical predictions for isotropic phase, bottom left, and nematic, upper right. Dotted lines are prediction of Onsager second virial approximation. Adapted from [46].

CHAPTER 2

Prebiotic scenarios

In the last decades, many attempts were undertaken (and some significant advances were obtained) to clarify some of the critical steps in life's origin and evolution, such as the synthesis of first building blocks, the origin of RNA and DNA or the first cellular organization [6]. However, also given the difficulty to verify some of environmental conditions on the early Earth, many of the issues are still highly debated [49], including the very definition of life. A definition that has attracted some consensus is the one proposed in 1994 by G. Joyce, and later adopted by NASA "Life is a chemical system capable of undergoing Darwinian evolution". A more explicit definition would include the ability of autonomous replication and the possibility to keep and propagate information (and thus to take advantage on natural selection) [50]. Quite different approaches are also proposed. An interesting one is by A. Pross [51], who proposes to focus less on history and developmental process of the species and assign the notion of "life" to individual entities capable of goal-driven actions.

Although the various definitions of life could in principle lead to identify "the" origin of life in different moments of the Earth history, in practice any definition points to what happened in an interval of about 500 millions of years, between around 4 billion years ago.

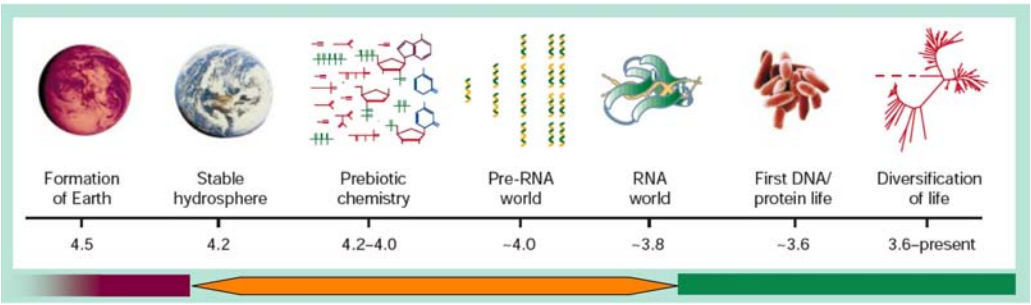


Figure 2.1: Timeline, expressed in units of billions of years, of Early life events, adapted from [5]. Colored bars indicate the era when the Earth could not have hosted life (purple), the era when life was certainly present (green), and the 500 million years interval (orange) when the origin of life took place.

In Figure 2.1, adapted from [5], the timeline of the main events regarding the origin of life is reported. The time axis can be thought as divided into two main sections,

as indicated by colors highlighting. The most recent portion (purple) of the timeline encompasses the range of time in which we have paleontological evidence of life on our planet. At the other extreme, there is a time interval (green) where no life could have been present because of the planet conditions. This leaves a gap where somehow the inanimate became animated. This is when crucial events took place and where OL research focuses its efforts. Possibly, the knowledge of the events in such time interval could even enable a better definition of life. Scenarios about the events in the OL time interval are formulated (i) either moving forward in time on the basis of the planetary and chemical condition of the early Earth in an effort to understand how complexity could have possibly formed, (ii) or moving backward in time on the basis of life as it is known nowadays and with the aid of geological and paleontological evidence, in an effort to identify the simplest and more ancient form of life.

Efforts to move "forward" in time need to be based on the necessarily partial knowledge of the early planetary conditions. The simplest organic compounds may have been present as soon as the Earth surface was filled enough by seas of water, the earliest evidence of crustal water being of about 4.3 Gyr ago. Hence in the range $4.3 - 4.0$ Gyr ago we may assume simple organic chemistry to have started being present, while, as detailed below, at $3.8 - 3.5$ Gyr ago some primordial form of life was present. In between there is a gap of 500 million years that also includes the so-called "heavy late bombardment" at ca. 3.9 Gyr ago, a set of collisions probably bad enough to sterilize any form of life.

In 1952, Stanley Miller (under the direction of Harold Urey) tested the possibility to synthesize organic compound from inorganic precursors. Indeed, by applying for days an electric discharge to mixed vapors of H_2O, H_2, CH_4, NH_3, at that time considered the most likely components of early Earth atmosphere, he obtained various organic chemistry molecules, including aminoacids, nucleobases (adenine, guanine) and fatty acids, some of the building blocks for biotic molecules [3]. Experiments later performed, by Miller himself, in more correct atmospheric conditions, didn't produce such a large variety of simple compounds [52]. However, it has been recognized that similar composition of simple organic compounds are found or produced in various conditions [4, 53].

This set of results indicate that, even if we cannot really tell which way it happened, the early Earth could have been generally, or locally, rich in simple organic molecules, not dissimilar from the basic building blocks of nucleotides, peptides, hydrocarbons. This is why Miller's intuition, even if based on a wrong assumption, turned out to be overall rather well confirmed. However, the availability of biomolecular precursors is far from indicating a path for the emergence of life. This becomes more evident by investigating the possible nature of the first and simplest forms of life.

2.1 The minimal form of life

Paelonthological evidence of life, such as fossilized bacteria, stromatolites, oxygen bearing minerals, date at least to 3.5 Gyr (billions of years) ago, and maybe to 3.8 Gyr [54]. How could these first forms of life be organized? An interesting approach to the problem is to find the "minimal gene set" from today's bacteria, i.e. the minimal ensemble

of genes that enable a bacterium to survive in some standard conditions. Experiments indicate that a set of about 80 genes is indispensable for a bacterium to survive [55]. The set contains the code for proteins devoted to transcription, translation, DNA replication, metabolism, cell division. This finding implies a rather sophisticated cell life, way too sophisticated to have emerged through a discontinuous process. Can this cell organization be further simplified?

The most convincing answer so far conceived to this question leads to the so-called "RNA world". RNA molecule has a pervasive role in contemporary biology, especially with regard to the most fundamental and highly conserved cellular processes. It is involved as a primer in DNA replication and as a messenger that carries genetic information to the translation machinery. Even more interestingly, RNA is a crucial component of the ribosome – the actuator of the translation. Hence, if DNA were replaced by RNA, the transcription and translation processed could be replaced by a straight translation of the genetic code into proteins. Furthermore, it has been found that RNA may structure in "ribozymes", i.e. RNA-made enzymes, that perform various catalytic activities, such as assisting in RNA processing events and in functions related to the replication of viral genomes. This evidence makes it reasonable to imagine RNA molecules capable to replicate themselves. If we could find an RNA polymerase (i.e., an enzyme promoting polymerization) that was itself a ribozyme, then a simple ensemble of molecules might be capable of self-replication. The protein-nucleic acid world of contemporary biology could have emerged later the course of evolution.

Although this sequence length is short with respect to the length of the genetic sequences, it is actually impossible to imagine the formation of such a polymer on the basis of random chemistry. Indeed, how the simple Miller-type molecules could have combined yielding life is the key question of the origin of life.

2.2 The RNA world

Inspired by the notion that RNA is a molecule in principle capable of carrying and duplicating information and folding into chemically active secondary structures, many investigations have focused on developing ribozymes. Strategies of test-tube evolution have enabled obtaining several examples of ribozymes able to catalyze the template-directed joining of a oligonucleotide terminated 3′-hydroxyl to a oligonucleotide terminated 5′-triphosphate [56]. RNA sequences have been found that enable ligating up to 20 nucleobases, while their own length is of the order 200bp [57, 58]. Recently, a possible natural ribozyme with 3′,5′ ligase activity was reported [59], giving for the first time real experimental support to the RNA-world scenario.

Despite these successes, the RNA-world view is disputed for various reasons. Firstly, although ribose, phosphate, purines and pyrimidines may have been all available in prebiotic environment, their combination in RNA oligomers would have been a low yield synthesis because of the presence of the much larger amount of competing nucleotide analogues. Indeed, the nucleotides (and their analogues) may even have joined to form polymers, with a combinatorial mixture of 2′,5′, 3′,5′ and 5′,5′-phosphodiester linkages, a variable number of phosphates between the sugars, D and L-stereoisomers of the sug-

ars, and assorted modifications of the sugars, phosphates and bases. The self-replication mechanism had somehow to accommodate these compositional differences and select the "right" nucleic acids. In addition, only conveniently activated nucleotides can be ligated to a chain. Actually, the phosphorylation of mononucleotides and the synthesis of short oligomers was demonstrated in suitable extreme environmental conditions [60], but today the usual laboratory route is to use phosphorimidazolides of nucleosides or other activating groups [61], favoring polymerization, whose presence in prebiotic environment has not been proved.

Another class of objections raised against the RNA world hypothesis pertains to the activities of RNA catalysts, i.e. to the mechanisms that must have led to the emergence of specific, rather long (despite the relative fragility of long RNA polymers in aqueous solutions), active sequences over all possible sequences. Indeed, although it was demonstrated that oligo(C)s as short as four monomer units in length can serve as efficient templates for the synthesis of oligo(G)s from activated monomers [62], a RNA fragment length of 50-100 is assumed to be required for a good catalytic activity. However, 50-mers could be assembled in approximately 10^{30} different sequences, corresponding, if one molecule per sequence is considered, to about $3.5 \cdot 10^7$ kg RNA, a small fraction of which with catalytic functions. This impressive compositional redundancy makes the emergence of functional sequences quite a challenge.

In summary, if the building blocks of RNA were available in the prebiotic environment, if these combined to form polynucleotides, and if some of the polynucleotides began to self-replicate, then the RNA world may have emerged as the first form of life on Earth. Assuming its validity, the RNA-world somehow solves the "chicken or egg" problem between nucleic acids and proteins, but still leaves the following question unanswered: how did the first poly-nucleotides arise from monomers, without any enzyme, of whatever nature?

2.3 Relevant conditions for nucleic acids formation

The critical step for the dawn of RNA (or DNA) as the information carrier lies in its elongation from single nucleotides, or oligomers, to the long polymer we know today. RNA self-replication cannot be achieved by one single molecule - at least two are needed - and generally for any workable chemical system one needs local concentrations of at least femto-moles of sequences and larger for the mononucleotides. Clearly, a relatively high concentration is a key requirement for the polymerization to take place and not to be overcome by chain degradation. High concentration alone is not sufficient, though, to sustain polymerization: like most of the present enzymes work by geometrical, physical constraint - by keeping close together active groups and thus enhancing reaction rates -, a template mechanism is invoked in prebiotic times to favor ligation of different nucleotides.

2.3.1 Building blocks concentration

To enable encounters and reactions between the simple molecules available on the early Earth, some kind of mechanism was certainly at play, especially since there is evidence that prebiotic oceans were as dilute as contemporary ones. Life is now organized into cells, very complex "worlds" basically separating genetic material (and various degrees of organelles and molecular machineries) from the outside through selectively permeable lipid membranes. The most natural and conservative approach would appear to imagine some simple proto-cells, combinations of RNA and surfactants, achieving the same results and able to replicate [63]. This possibility is made more interesting by the fact that some surfactant micelles and vesicles were found to spontaneously split (and thus "self-replicate") under appropriate conditions, as schematized in Figure 2.2 [64, 65]. Although some similar phenomenon had certainly to occur for the birth of the first cell to take place, it appears unlikely that this was the real driving force for the original RNA segregation and elongation, since this would imply a locking mechanism between the duplication of vesicles and the evolution of RNA, which would in turn imply sophisticated machineries reflecting, through molecular synthesis, the RNA sequence to the behavior of the vesicles.

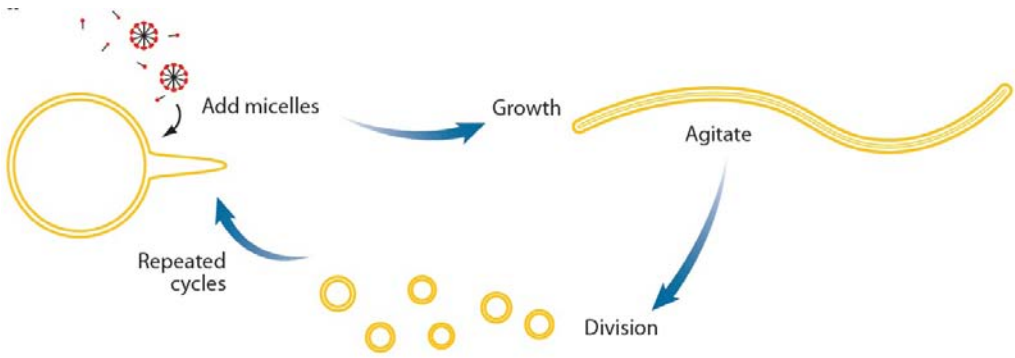


Figure 2.2: Model for protocells replication from fatty acid vesicles growth and replication, from [65].

Hydrotermal marine environments, characterized by heat currents flowing through porous minerals, could have played a role in the development of life, providing a heat source, minerals in solution and fluctuating conditions. A spontaneously increased concentration of RNA strands could have been promoted by convection and thermo-diffusion [66], see Figure 2.3. Recently in such systems hyperexponential escalation of polymerization of > 100 bases long RNA strands has been achieved [67]. In a rough estimate, a 10^6-fold accumulation is required for small protobiomolecules to interact. Also surfaces and structured porous minerals could have promoted, by preferential adsorption, an increased surface concentration of prebiotic molecules and act, at the same time, as catalyzers, as discussed in the next section.

All these mechanisms lack, however, of the necessary capacity of selective concentration. Enclosure into vesicles and thermo-diffusion could have promoted local enhance-

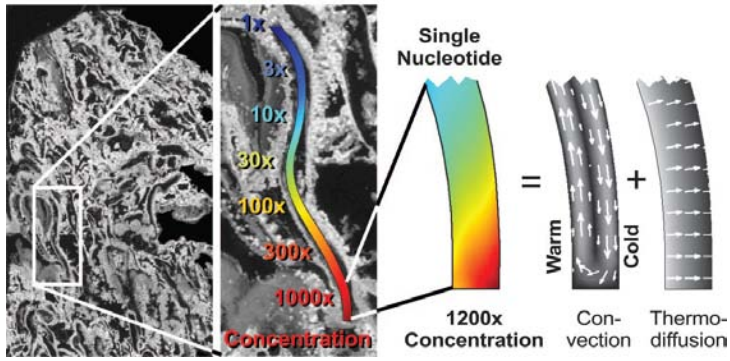


Figure 2.3: A spontaneously increased concentration of RNA strands could have been promoted by convection and thermo-diffusion in hydrothermal phores. Figure from [66].

ment of molecular species, but without relevant selectivity among the huge ensemble of molecular variants. Surface adsorption is more specific, but it is hard to imagine an adhesion process that would be strong enough to induce local crowding but weak enough to enable mixing, collisions, interactions and all the molecular events necessary to the formation of complexity.

2.3.2 Template

The appeal of RNA as the first self-replicating molecule relies on the fact that, by definition, it would be capable of acting autocatalytically for its own synthesis and, at the same time, such autocatalytic molecule would act as a template to bind the precursors by non-covalent forces and organize them in such a way that the reactive groups come in close proximity (see scheme in Fig. 2.4). Studies with activated trimers and hexamers showed that template autocatalysis can only occur if the sequences of both trimers match the sequence of the hexamer according to the Watson-Crick base-pairing rules. They also showed that the condensation reactions are predominantly controlled by the stacking of nucleic acid bases flanking the newly formed internucleotide link [68]. As already said, however, we are left with a new problem: how did the first templating RNA oligomer arise, without a template?

Bulk condensation polymerization reactions are usually thermodynamically driven towards hydrolysis in dilute aqueous solutions. Therefore, besides high concentration, a surface-promoted mechanism is required to enhance the polymerization rates. Some mineral surfaces have been proposed as good templates for nucleotide polymerization. The most credited candidate is montmorillonite, a clay mineral with a layered structure. Reversible hydration or solvation of the cations cause the layers to expand, favoring the entrance of certain molecules [69]. A number of experiments verified the binding of mononucleotides onto the montmorillonite surface or inside its layers and its ability to promote the formation of the phosphodiester bond (in suitably activated monomers) and thus the elongation of nucleotide polymers [70, 71]. Interestingly, montmorillonite was also reported to favor the homo-chiral selection of nucleotides [72], another critical

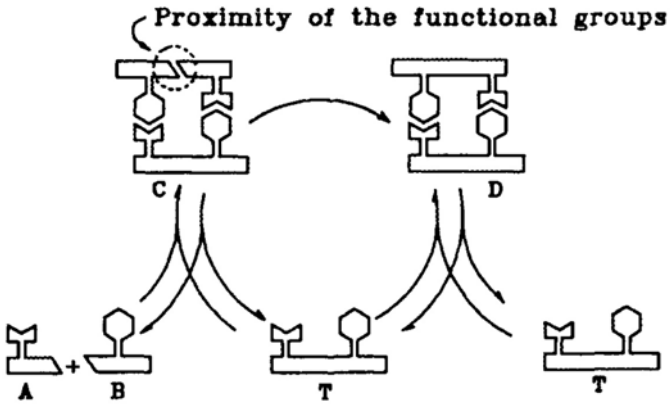


Figure 2.4: Schematic representation of self-replicating system: a template T catalyzes its reproduction trough the ligation of two nicked complementary sequences A and B, from [68].

step in the development of longer molecules.

However, in clays the catalytic surface is a liquid-solid interface, and the lack of fluidity in this interface could be poorly compatible with an efficient surface diffusion [73]. Such surfaces also lack the flexibility which is found in present-day enzymes, and which is known to be crucial to catalysis. Accordingly, some other mechanisms have been proposed, involving liquid or "soft" phases. Oparin [74] suggested that prebiotic polymerization reactions took place in a heterogeneous, coacervated system, rather than in the bulk of a homogeneous phase. Coacervation, a liquid-liquid phase separation, was considered as an essential concentrating process by which mixtures of randomly formed prebiotic polymers initially in dilute solutions were condensed into concentrated assemblies. Although naive (he thought that coacervated droplets directly lead to cells), Oparin's theory introduced for the first time the idea that a physical phase separation process could lead at least to locally enhanced concentration of nucleotides and thus more favorable elongation.

The proposal of the catalytic role of a liquid-liquid interface, namely between an oil slick and salty water, was put forward by Lars Onsager [75], and the renaturation process of DNA oligomers is found to be enhanced at the interface between phenol and water [73].

Recently [76], RNA monomers were found to polymerize when subjected to dehydration cycles, at moderate temperature, in a mixture with simple lipids. At high concentrations, the lipids organize in lamellar and possibly hexagonal phases, thus providing a template for the linear elongation of the RNA strand. Again, the observed spontaneous LC ordering of short RNA strands emerges, in our opinion, as the most simple and non-redundant template mechanism: without relying on additional ingredients, it provides a self-templating structure. The LC matrix provides a flexible template, possibly favoring chemical ligation, and thus promoting an auto-selective process, since longer helices can more easily fit in the aligned environment.

Some very recent studies show the possibility to achieve spontaneous ligation of nucleic acids monomers or oligomers in a quite extended range of moderately high temperatures and pH [77] under the combined effect of sequence complementarity and high concentration [7].

2.4 DNA self-assembly as a prebiotic route

Nowadays leading scientists in the origin of life field invoke self-assembly mechanisms to tackle key open questions of the origin of life. In this perspective the phenomena involved in LC ordering of DNA duplexes could provide in a single shot a number of distinctive features: (i) a mechanism of selection, since only paired strands segregate to form LCs; (ii) a locally enhanced concentration of complementary strands and proximity of terminals (iii) a liquid crystalline matrix which provides a flexible template, favoring the linear aggregation and possibly the chemical ligation; (iv) an autocatalytic process by which longer complementary sequences are favored, because they can more easily fit within the ordered environment.

This conjecture and our recent results in testing its reliability are widely discussed in Chapter 6.

CHAPTER 3

Materials and experimental techniques

3.1 DNA synthesis

The largest part of oligoDNA was synthesized using an Äkta Oligopilot (Figure 3.1), manufactured by Amersham Pharmacia Biotech. The majority of coupling reagents were ordered from Sigma-Aldrich. Amidites, solid support (nucleotide bound on polystyrene), and ethylthio tetrazole were purchased from American International Chemical, Inc. For random oligonucleotides, randomness of the synthesized oligomers at the designed positions in the sequences was obtained by providing the oligopilot with a single vessel containing an equimolar solution of each of the four phosphoramidites. For defined oligomers 4 separate vessel, one for each of the four phosphoramidites were normally adopted. Phosphoramidites were purchased and utilized only if both coupling efficiencies and purity level were greater than 99%, to ensure no reactivity differences. Given equivalently high levels of both purity and reactivity for each individual phosphoramidite, it is safe to conclude that no significant bias toward any specific base is expected in the syntheses of oligomers when the oligopilot is provided an equimolar solution of all of the phosphoramidites.

In the last step of synthesis, we omitted removal of the 4,4'-dimethoxytrityl (DMT) blocking group to facilitate later purification. Materials synthesized using the Äkta Oligopilot came off the synthesizer as bound oligoDNA on solid support. OligoDNA was cleaved from the solid support by stirring the solid support-bound oligoDNA in an ammonium hydroxide solution (ammonia content approximately $28 - 30\%$) in a ratio of 40 ml per 1.6 g of solid support material. The suspension was heated to 50° Celsius for 8 hours, then allowed to cool to room temperature. The cooled suspension was vacuum filtered over a coarse frit and the solid filtrate washed with three portions of an ethanol/water solution (1:1 by volume, total volume of $400ml$). The resulting solution was then stripped of solvent under vacuum to provide the crude unpurified oligoDNA.

Crude oligoDNA was purified using preparatory HPLC. A dual Varian Prostar 218 system tied to a Prostar 318 detector employing a SepaxGP-C8 column from Sepax Technologies Inc was used to remove failure sequences from the crude oligoDNA to yield purified random 20meric oligoDNA. The oligoDNA was purified by the "DMT-on" HPLC purification method. Crude oligoDNA was suspended in 50 mM triethylamine bicarbonate (TEAB) buffer with 100 mM sodium chloride and filtered through a $0.2\mu m$ nylon

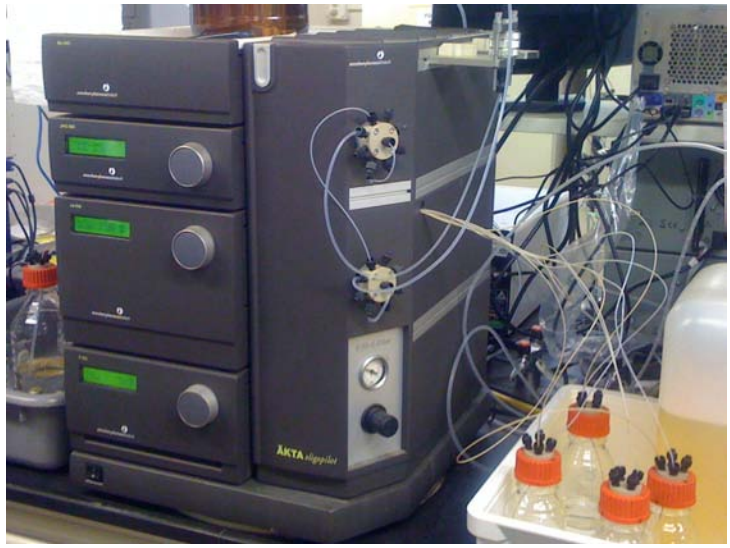
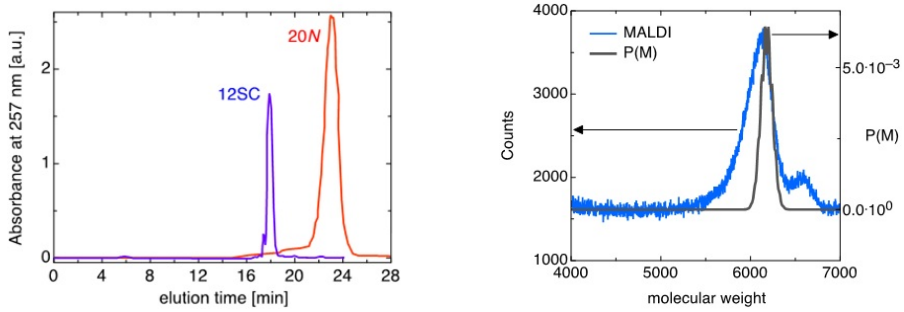


Figure 3.1: Picture of Akta Oligopilot used for the synthesis of DNA oligomers. The four glass vessels on the right are the phosphoramidites containers. Cylindrical reaction chamber containing the solid support is visible at the upper-right side of the Oligopilot.

syringe filter to remove particulates. OligoDNA was then passed through the HPLC using 50 mM TEAB as the primary buffer and HPLC grade Acetonitrile as the secondary buffer. DMT-capped oligoDNA separates from failure sequences on the column using a slow linear elution gradient with increasing Acetonitrile concentration. The presence of the DMT-cap causes the sequence of interest to elute much later than the uncapped contaminating sequences, giving efficient separation. Once eluted from the HPLC column, the polynucleotide was concentrated to dryness using a rotary evaporator. 40 ml of 80% Glacial acetic acid with 20% water was added to the dried oligoDNA and mixed one half hour at room temperature, with shaking, to cleave the DMT group from the 5′-end of the polynucleotide. Acetic acid was predominantly removed by evaporation on the rotary evaporator and then the oligoDNA was re-suspended in 40 ml of 50 mM TEAB with 2% ammonium hydroxide by sonicating at room temperature for several minutes. A syringe filter was again applied to remove particulates so that the solution could be sent through the HPLC a second time using the same column, buffers and elution gradient as the initial run. In the second run, the de-protected oligoDNA elutes from the column at the same hydrophobicity as the failure sequences eluted in the initial run. Following elution from the column, the oligoDNA was concentrated to dryness using the rotary evaporator. At this step, the purified DNA can be dissolved in water and used as necessary for other experimentation. MALDI-TOF mass spectrometry was used after each HPLC step to confirm the presence of the expected oligoDNA.

Parts of the sequences – HPLC purified – were purchased from PRIMM (Segrate, Italy) as liophylized powders.



(a) (b)

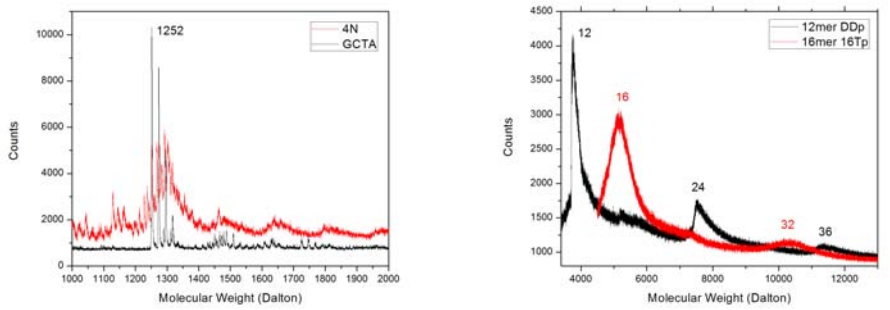
Figure 3.2: (a) HPLC traces for the synthesized compound 12SC and 20N. (b) MALDI-TOF mass spectrum for the 20N (blue curve, left axis) compared with the calculated normalized mass distribution $P(M)$ (black line, right axis).

3.2 DNA characterization by HPLC and MALDI mass spectroscopy

Initial characterization was performed by HPLC. We show in Figure 3.2a the HPLC traces obtained with a single distinct 12mer DNA (CGCGAATTCGCG, Dickerson Dodecamer, 12SC) and of the fully random 20mer (20N). As compared to the purification of 12SC, HPLC traces of the final step of 20N purification have a broader elution profile which displays characteristics of a composite peak. This would be expected of a sample containing many different DNA sequences of a characteristic length, but each eluting at a slightly different condition of hydrophobicity due to the variations in A,T,C, and G content. It also eluted somewhat later than the 12SC, suggesting a uniformity of elution depending on the number of bases present in all of the 20N sequences.

The 20N DNA was further characterized by MALDI-TOF mass spectrometry as shown in Figure 3.2b (left axis). In order to estimate the width of the expected mass distribution for 20N, we generated 10^8 random 20mer sequences on the computer using the Random Number Generator by Wolfram Mathematica and determined their expected mass distribution $P(M)$, shown in Figure 3.2b (right axis). The overplot on the Figure enables comparison between the measured and the calculated mass distribution. The MALDI peak was approximately 6110 MW as compared to the calculated theoretical average weight of 6160 MW. The added broadness at the lower molecular weight fringe of the peak is likely more populated than the theoretical average as a result of some small-scale damage to the polymer done in the process of laser ablation. The purified 20N showed an additional contaminant peak at slightly higher molecular weight suggesting the inclusion of some 20N still containing the 5'-terminal protecting group left appended for purification.

MALDI-TOF mass spectrometry was used also to characterize other produced sequences and to identify the products of ligation reaction experiments (see Chapter 6. In Figure 3.3a we report MALDI-TOF spectra of random 4mer (4N) and definite sequence GCTA: the weight distribution of 4N, containing $4^4 = 256$ different sequences, is broader



(a) (b)

Figure 3.3: (a) MALDI-TOF spectra: characterization of random 4mer (4N) and definite sequence GCTA; (b) identification of reaction products of non-enzymatic ligation with 12mer DDp and 16mer 16Tp building blocks.

than that of the defined sequence GCTAp. The MALDI peak for GCTA was at 1252 MW as compared with the calculated weight of 1253.8 MW. In Figure 3.3b we report the identification of reaction products of non-enzymatic ligation with 12mer DDp and 16mer 16Tp building blocks. MALDI-TOF measurements have been compared with gel electrophoresis data in order to double check the identification of ligation products.

3.3 Ligation reactions with EDC

Ligation reactions (see Chapter 6) were performed in microscope cells, Eppendorf tubes or glass capillaries. Here the followed protocol is reported.

3.3.1 Microscope Cell

Typically 1-5 OD of lyophilized DNA was dissolved into $3-15\mu l$ of freshly prepared 3-9 mM EDC solution in HEPES buffer (0.01M, pH 7.5), DNA concentration was then 3mM (ISO). $1\mu l$ of stock solution was loaded on a microscope glass plate and evaporated (5 - 15 min) up to the required concentration for LC phases (90-200 mM), then closed with cover slip and sealed with a fluorinated oil, in order to guarantee both concentration and thermal stability to the sample. At the same time $1\mu l$ of the stock solution was withdrawn and the reaction was arrested by dilution with $0.2ml$ of 50mM Ethanol Ammine, in order to obtain a "zero point" for the reaction. Both microscope cell and stock solution tube were stored at the same temperature of $25°C$ for the whole reaction time, 2-48 h (mostly for 24h). Then microscope cell was opened and reaction was stopped by dilution with $0.2-1ml$ of 50mM Ethanolamine, similarly stock solution was diluted 100 times with 50mM Ethanolamine. Final solutions were stored at $-20°C$.

3.3.2 Eppendorf tube

For concentration scales the required DNA concentrations 3-120 mM were obtained in $1-15\mu l$ of equimolar solution of EDC in 0.01-0.1 M HEPES buffer, directly into Eppendorf tubes, reacted at $4-25^{\circ}C$ for 24h, then reaction was stopped by dilution with $0.2ml$ of 50mM Ethanolamine and stored at $-20^{\circ}C$.

For experiment with PEG or single strands, the right amount of lyophilized oligomers (20-50 OD) was hydrated with PEG or single strands solution containing fresh prepared EDC solution in pH 7.55, 0.1M HEPES buffer. Samples were sealed by siliconic oil to prevent evaporation. Reaction was stopped by 20 fold dilution in 50mM Ethanolamine after 24 hours reaction time.

3.3.3 Capillary

The required quantity of DNA (20-50 OD) was lyophilized inside a bottom closed glass capillary ($2mm$ diameter), $0.5-1\mu l$ of 0.01M HEPES was added up to reach the concentration for LC phases (90-200 mM). Samples were sealed by siliconic oil to prevent evaporation. Generally capillary were heated at $70^{\circ}C$ and spinned to let the DNA-LC collect at the bottom and its concentration homogenize. Reaction conditions and arrest as previously shown.

3.3.4 Radio end-labeling of DNA oligomers

For quantitative yields measurements, radio end-labeling of 2 pMol of DNA products of ligation reactions, with $\gamma-[32P]-ATP$ (Perkin Elmer), was performed via T4 Polynucleotide Kinase (Epicentre Biotechnologies) in $20\mu l$ reaction volume following standard protocol. Already 5' phosphorilated oligonucleotides were previously dephosphorylated using FastAP Thermosensitive Alkaline Phospatase (Thermo Scientific), following standard protocol for $20\mu l$ reaction. In order to remove unconverted $\gamma-[32P]-ATP$, radio labeled DNA strands were purified with size exclusion Micro Bio-Spin Chromatography Columns with Bio-Gel P-6 gel (Biorad).

3.4 Polyacrylamide gel electrophoresis

Vertical polyacrylamide gel electrophoresis apparatur (Biorad protean II) was used for oligomers size separation (Figure 3.4). Dried oligonucleotides were hydrated with $2\mu l$ of denaturing gel loading solution ($\%30$ glycerol, 0.025%bromophenol blue, 0.25% xylene cyanol, water) and $10\mu l$ formamide, loaded into 15% denaturing polyacrylamide gel ($0.75-1mm$ thick), 7M urea, TBE buffer and run for 1-4 h at $30V/cm$ voltage. For radio labeled oligomers, gels were viewed on Personal Molecular Imager FX System (Biorad) after 5 to 90 min exposure on radiographic plates. For non radio end-labeled sequences, $100-500ng$ of DNA was loaded and, after the run, gel was stained for 15 min in $0.5\mu g/ml$ EtBr in TBE buffer. Gels were viewed on UV-vis illuminator and images acquired with Typhoon-9200 Phosphor Imager (Amersham Pharmacia Biotech). To

estimate products lengths a proper ladder was made with 12mer, 24mer, 36mer, 48mer oligomers synthesized by repetitions of starting DD sequence (Primm).

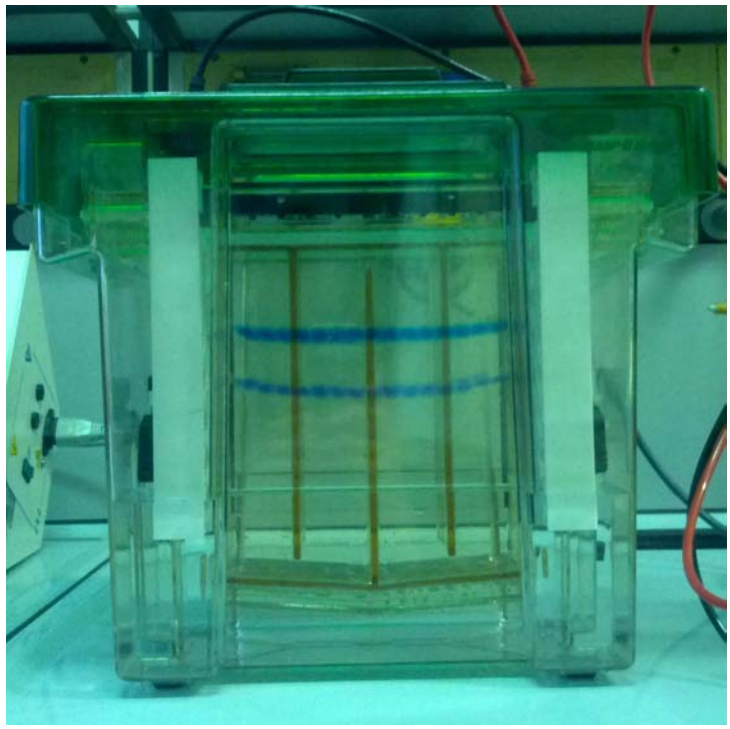


Figure 3.4: Picture of gel electrophoresis apparatus Biorad Protean II, gel size $16x16cm$.

3.4.1 Analysis of gel lanes

B/w 8bit acquired gels images were processed with MacBiophotonics ImageJ software. Brightness and contrast adjustment were made in order to maximize the signal to background ratio, paying attention to don't saturate the image. Successively the profile of each lanes was plotted. Profile plots were acquired with Origin software that allowed the measurements of peaks positions and the integral of each peak. In Figure 3.5 we report an example of analysis of an EtBr stained gel.

Assuming linearity between the area of the peaks and the DNA contents it is possible to measure the amount of products of each gel band. For $\gamma - [32P]$ labeled strands this assumption is peaceful since each marked strand has one radioactive Phosphate, than the radio signal is directly proportional to the number of strands, i.e. in order to obtain the number of monomers the signal has to be multiplied for the degree of polymerization.

For EtBr stained gels, instead, since the amount of EtBr intercalated in each strand is not controllable, the fluorescence signal is not an absolute measurement of the number of strands. Nevertheless, assuming that EtBr intercalate in the same way for each DNA strand of the same length, it is possible to obtain a relative measurement of DNA mass

in each gel band. This assumption is sustained by the analysis of ladders bands in which the DNA amount was known a priori. For example in Figure 3.5d, we report the plot of the peaks integral for two ladders made of 12mer, 24mer, 36mer and 48mer at the same molarity. Since each sequence is obtained by the repetition of 1, 2, 3, 4 times the sequence of the starting 12mer, the mass of each peak should be equal to the mass of the first peak multiplied for the degree of polymerization. The plot shows that the peaks integral is not so far from this predicted behaviour and linearity is maintained at least for the first tree points, while DNA mass of the forth peak is underestimated.

Therefore despite EtBr staining doesn't allow a absolute quantitative measurements of the DNA content of a gel band, it allows a relative measurement and is a good estimation of the mass distribution.

Moreover peaks position allows the identification of the products length, following the empirical low that electrophoretic mobility of a DNA strand is proportional to the logarithm of the number of bases, as shown in Figure 3.5c.

3.5 Optical microscopy

Microscopic observations of cells and capillaries were performed either on a Nikon TE200, equipped with a fluorescence filter set, or on a Nikon Eclipse Ti phase contrast microscope, both linearly polarized. Images were acquired on a Nikon DS-5M color or on a Prosilica GX-1050 b/w CCD cameras.

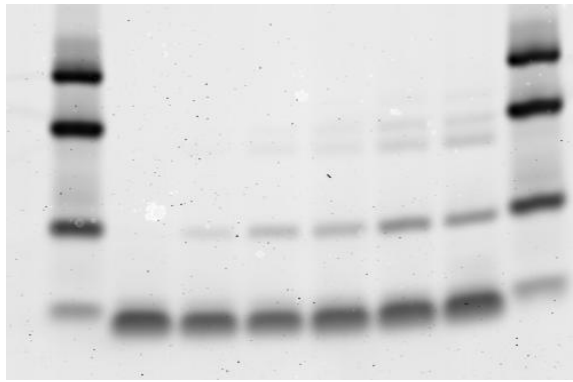
3.5.1 LC observation

The optical properties of liquid crystal phases often directly reflect the symmetry of their structures. Birefringence, anisotropy of the refractive index, allows the visualization of the macroscopic molecular orientation. A simple liquid is isotropic and its refractive index is independent on the direction of the incident light beam, thus if placed between two crossed polarizers no light will be transmitted and the sample will appear black. Instead in thin liquid crystal sample cells placed between two crossed polarizers under an optical microscope, various textures and birefringence colors will be observed. These textures and colors not only look beautiful but also provide a lot of information about the structure of the LC phases. Although there are many experimental techniques to investigate the structure and physical properties of LC phases, microscope observations often give enough information to determine the structure.

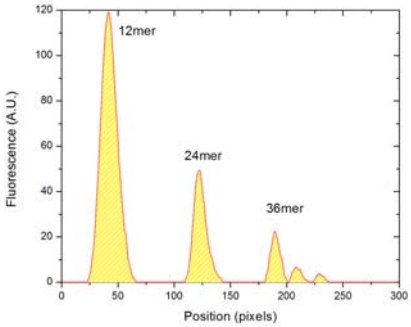
Textures of the LC phases of interest in the next Chapters are: Blue Phase (BP*), Chiral Nematic phase or Cholesteric (N*), Columnar Uniaxial (C_u) and high concentration Columnar (C_2). Exhaustive study of LC phases of short DNA oligomers has been performed in [78]. Other LC textures and detailed explanations about phase characterization from polarized microscopy can be found in [79].

3.5.2 Concentration measurements via microscope-based interferometry

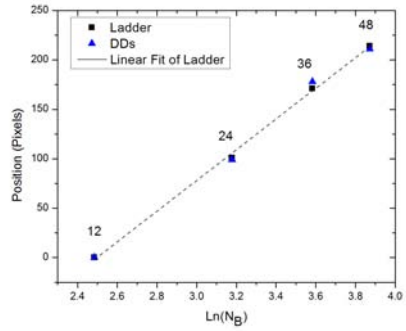
The measurement of DNA concentration in sealed cells was performed by microscope-based interferometry. Planar cells made with high-refractive index glass (F2 Schott,



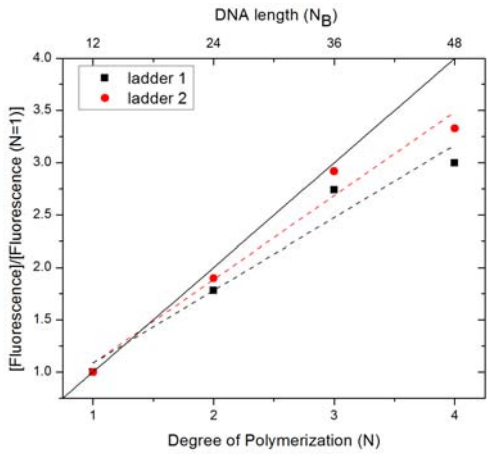
(a)



(b)



(c)



(d)

Figure 3.5: Example of analysis of EtBr stained gel (a): (b) plot of the intensity profile of the second to last lane of Figure (a), background subtraction has been applied. (d) Identification of oligomers length by plotting the final bands position vs $\ln N_B$. (d) Plot of the integral of the intensity profile of the two ladders in gel (a) divided by the integral of the 12mer peak, solid line is the predicted behaviour $y = x$, dashed lines are linear fits.

$n = 1.62$) were brought, through heating, into the isotropic phase and the spectrum of the light reflected by the illuminated portion of the cell was collected. The multiple reflection fringes due to the interference from the parallel glass plates were systematically measured in various spots within the DNA solution and on the periphery of the cells, where the fluorinated oil was instead present. This set of data enabled determining the refractive index of the DNA solution and thus the DNA concentration c_{DNA} through a calibration curve ($n = n_0 + dn/dc \cdot c$), with n_0 the refractive index of the solution at zero DNA concentration and $dn/dc = 0.136 cm^3/g$). The extracted values have an uncertainty of about 5%.

Concentration of diluted stock solutions were measured with NanoDrop 2000 (Thermo Scientific).

3.5.3 Melting Temperature Measurements

The duplex denaturation transition was probed by measuring the fluorescent emission of Ethidium Bromide (EtBr) mixed into DNA solutions at a concentration of one molecule per duplex. Upon slowly increasing T, as the duplexes unbind, the EtBr fluorescence is greatly reduced and eventually reaches a hight-T plateau, enabling the determination of the fraction of paired strands. The melting temperature is defined as that corresponding to half of the population of unpaired helices.

3.6 Computer-based estimates of the statistical properties of ranDNA

3.6.1 "Nearest neighbor" (NN) model

DNA double helices are formed by two distinct DNA strands held together by the combination of pairing and base stacking forces. These forces control the selectivity of the binding process, by which the binding energy depends on the degree of complementarity of the nucleobase sequences along the two polymers. Base pairing originates from the formation of hydrogen (H) bonds between pairs of nucleotides. The strength of the interaction is maximum when the sequences match according to the Watson-Crick (WC) complementary couples, i.e. adenine (A) with thymine (T) (two H bonds) and guanine (G) with cytosine (C) (three H bonds). The energy associated with the formation of the double helix ("hybridization"), depends on the length of the nucleic acid polymer and on the quality of the WC matching, vanishing for pairs of sequences with poor complementarity. Base stacking forces, instead, are attractive forces between the aromatic hydrocarbon plates of the nucleobases. Stacking is much less selective than pairing and its strength mainly depends on the overlap between the aromatic plates.

The most commonly adopted approach to evaluate and predict the overall binding free energy ΔG between hybridized DNA strands relies on the so-called "nearest neighbor" (NN) model [11, 12]. According to it, the double helix is decomposed in quadruplets formed by two consecutive nucleotides on one strand and the corresponding nucleotides on the other strand. This is shown in Figure 3.6a, two partially complementary 10mers are taken as an example. 10 quadruplets can be distinguished, that include also the two "triplets at the extremes". Notice that the quadruplets overlap. The NN model neglects

contribution arising from non-local (beyond nearest neighbors) interactions along the helix. According to this model, the total free energy involved in the formation of the double helix can be estimated by

$$\Delta G = \sum \Delta G_Q + \Delta G_{INIT} \tag{3.1}$$

where the summation is on all the quadruplets of the sequence, ΔG_Q is the free energy of each quadruplet and ΔG_{INIT} is an "initiation" free energy. The analysis of the database in [12] indicates that ΔG_Q significantly promotes the formation of the double helix only when the quadruplet is formed by two WC pairs. When averaged over all the possible quadruplets involving WC pairs (AA/TT, AT/TA, TA/AT, CA/GT, GT/CA, CT/GA, GA/CT, CG/GC, GC/CG, and GG/CC), $\langle \Delta G_Q \rangle \sim -1.6 kcal/mol$ at $T = 25°C$, corresponding to $\sim 2.7 k_B T$. Pairing errors in one of the two couples lead, on average, to an increased free energy: $\langle \Delta G_Q \rangle \sim 0.4 kcal/mol$.

3.6.2 Adopted assumption

The computer based free energy evaluation reported in this study (and used beyond in Chapter 4) are based on the NN model in which further simplified by the following assumptions:

1. we neglect differences in free energy contributions among quadruplets. Accordingly, quadruplets 3, 4, 5, 8 and 9 all give the same contribution $\Delta G_0 = \langle \Delta G_Q \rangle$ to the binding free energy, marked, in Figure 3.6b with staple shapes below the duplex;

2. we neglect the contribution to the free energy of quadruplets in which one or both pairs aren't WC pairs. Thus, the contribution of quadruplets 2, 6 and 7 to the duplex binding free energy is set to 0;

3. we neglect the contribution to the duplex stability of "coaxial stacking", the situation occurring when one base is overhanging at the end of a duplex (triplets 1 and 10), always quite smaller than ΔG_0;

4. we neglect ΔG_{INIT}. Among the approximations here introduced, this one is the most significant. A better approximation would have been to set $\Delta G_{INIT} = \Delta G_0$, as their values are comparable. Introducing this amelioration in the calculations brings to a shift of $-1 \Delta G_0$ in the free energy scales of Figure 4.12 and a different value for the attempt rate in the duplex lifetime equation, i.e. $\tau_0 \sim 150 ns$. However, we decided not to include this variant because it makes the discussion more complex without introducing any real change in the results. Furthermore, ΔG_{INIT} has been well investigated for duplexes with a large fraction of WC pairing, and it may not simply translate to situations as those explored here, where the degrees of freedom within the duplexes are much larger because of the high degree of pairing errors.

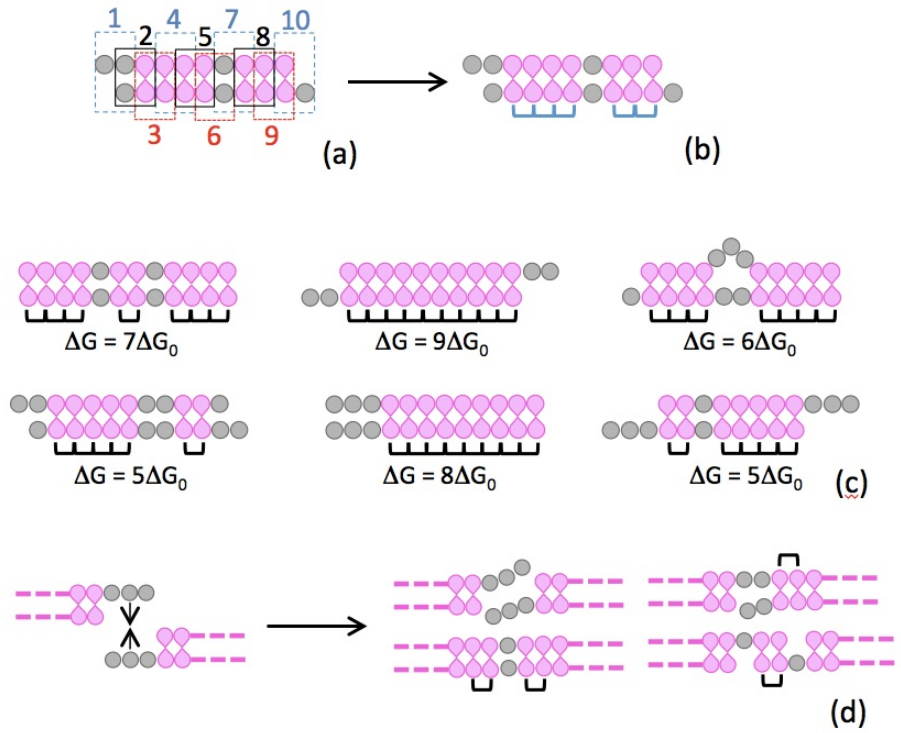


Figure 3.6: (a) Graphic description of the approach used to evaluate the binding free energy ΔG of hybridized DNA strands relies according to the nearest neighbor model. The double helix is decomposed in quadruplets formed by two consecutive nucleotides on one strand and the corresponding nucleotides on the other strand. Notice that the quadruplets overlap. (b) Out of the 10 quadruplets in the example reported in panel (a), only 5 are correctly paired (marked with blue horizontal squared brackets below the drawing. (c) Example of the application of the simplified NN approach to compute the energy of variously paired 10mers. (d) Graphical description of the inter-duplex interactions mediated by random overhangs. The simplified NN model enables evaluating the inter-duplex energy for overhangs with various complementarity.

3.6.3 Evaluation of intra-duplex binding free energy

These simplifying assumptions enable a prompt estimate of the duplex pairing free energy, as in the examples in Figure 3.6c. This simplified approach to the computation of duplex binding free energy has enables us to perform statistical evaluation on large numbers of computer generated partially paired duplexes, as described in the following.

3.6.4 Evaluation of the inter-duplex interaction free energy for DD with random overhangs

For each length R of random overhangs, 10^7 collisions of randomly generated overhangs were considered. For each, the value of free energy has been determined as described above. This process is sketched in Figure 3.6d. For $R > 2$ shifted configurations were

also included (as in the last example of Figure 3.6d). The resulting values for $\langle \Delta G \rangle$ are reported in Chapter 4 and in Figure 4.13 (right axis).

3.6.5 Evaluation of $n(\Delta G)$

Within the ensemble of fully random sequences of a given length, $n(\Delta G)$ expresses the number of possible oligomer combination yielding a given duplex binding free energy ΔG. As described in Chapter 4, random collisions between randomly generated sequences yield a large variety of pairing motifs, some of which are sketched in Figure 3.6c for fully random 10mers (10N). An important point is that by "random" collision we mean that we evaluate the energy not only for "in register" couples of oligomers (potentially leading to blunt-ended duplexes if terminals produce WC pairing), but also for oligomers that are brought in contact with a shift of arbitrary (randomly determined) length, a situation giving rise to duplexes with overhangs.

More specifically, for the case of 20N, $n(\Delta G)$, as reported in Figure 4.10, is expressed as the number of combination that one given 20mer sequences can make with the other ones of the 20N family so to have a binding strength equal to ΔG. Accordingly, $n(19\Delta G_0) = 1$, since in the 20N pool one only sequence is the perfect complementary of the starting sequence. Analogous definitions apply to the 6N and 12N cases reported in the Figures of Chapter 4.

$n(\Delta G)$ has been generated by merging its behavior at small and large ΔG, that were independently evaluated. By generating, at the computer, a number N_{SIM} of duplexes in the range $10^7 - 10^8$, and evaluating the binding free energy for each of them, have obtained the free energy distribution at low ΔG. This approach is necessarily limited to small ΔG because the number of sequences that can be generated is small compared to the $N_{TOT} = 4^{20} \approx 10^{12}$ sequences of the 20N system. In this small ΔG limit, $n(\Delta G)$ is obtained by multiplying the number of computer generated duplexes at each value of ΔG by the factor N_{TOT}/N_{SIM}, i.e. the ratio between the number of different sequences actually present in a rDNA system and the number of sequences generated by the computer. The behavior of $n(\Delta G)$ for large values of ΔG can instead be analytically determined. For example, in the case of 20N, by examining the possible source and location of errors in pairing, it can be recognized that $n(19\Delta G_0) = 1$, $n(18\Delta G_0) = 14$, $n(17\Delta G_0) = 161$, $n(16\Delta G_0) = 1681$, $n(15\Delta G_0) = 10970$. These values are the rightmost blue squares in Figure 4.10. As visible in the figure, despite the gap between the two approaches $n(13\Delta G_0)$ and $n(14\Delta G_0)$ couldn't be evaluated either analytically or with computer simulation), the two different determination of $n(\Delta G)$ very well match and are perfectly fitted by the product of an exponential decay and a Gaussian.

The same combination of approaches enabled us to determine $n(\Delta G)$ for the 6N and the 12N rDNA solutions reported in Figure 4.10.

3.6.6 Evaluation of the equilibrium and non-equilibrium distribution in duplex binding free energy $P(\Delta G)$

Equilibrium distribution are obtained by multiplying $n(\Delta G)$ and the Boltzmann factor $\exp(\Delta G/k_B T)$, both plotted in Figure 4.10. The resulting distributions are shown, after

normalization, in Figure 4.11.

Kinetic distributions are determined via computer simulation by following the time evolution one given sequence through the consecutive formation of duplexes with other randomly chosen strands. Specifically, the non-equilibrium distribution is obtained through the following steps:

1. one specific ("original") sequence is randomly chosen and held;

2. randomly chosen partner sequences are produced, and their binding energy with the original sequence are determined, as discussed above, for given randomly chosen shifts;

3. on the basis of the binding free energy of each of these randomly generated duplexes, we determine their lifetime as explained in the main text;

4. we sum the duplexes lifetimes and we keep generating random sequences to be duplex partners to the original one until the total lifetime reaches the desired value, chosen to approximate the characteristic experimental time (in this work we assumed $\tau_{TOT} = 10 hours$);

5. we generate a new original sequence and repeat the procedure. This approach enables to evaluate the population of double helices that are expected after a given time from the preparation of the rDNA solution and its cooling at room temperature. The resulting distributions are shown in Figure 4.12, where it can be appreciated that for the shortest sequence (6N), this kinetic approach leads to a distribution identical to the calculated equilibrium distribution. 12N is also almost equilibrated, while 20N appears far from equilibrium.

Non-equilibrium distributions were evaluated on the basis of about 50 original sequences.

CHAPTER 4

Liquid crystal self-assembly of random-sequence DNA oligomers

In biological systems and nanoscale assemblies, the self-association of DNA is typically studied and applied in the context of the evolved or directed design of base sequences that give complementary pairing, duplex formation, and specific structural motifs. Here we consider the collective behavior of DNA solutions in the distinctly different regime where DNA base sequences are chosen at random, or with varying degrees of randomness. In these systems the large body of different competing sequences effectively reduces the selectivity of the interactions.

We show that in solutions of completely random sequences, corresponding to a remarkably large number of different molecules, e.g., $\sim 10^{12}$ for random 20mers, complementary still emerges and, for a narrow range of oligomer lengths, leads to structured self assembly in the form of organization of the single strands into liquid crystal (LC) phases. This ordering follows from a subtle hierarchical self-assembly sequence in which the kinetic arrest of oligomer association into long-lived partially paired double helices is followed by reversible association of these pairs into linear aggregates that condense into LC domains.

With these results, we show that the phenomenology of the self-assembly of nucleic acid oligomers is actually much richer than previously recognized, involving self-selection, linear aggregation and ordering of fully random chains. Our results strengthen the notion that DNA and RNA have unequaled capacity of self-structuring and unavoidably suggests self-assembly as the possible key factor for the emergence of nucleic acids from the prebiotic molecular clutter as the coding molecules of life.

4.1 Random DNA oligomers

Random sequence DNA is an important class of systems to understand in regard of the appearance of early life, as pools of heterobase oligomers are likely prebiotic systems emerging from random ligation. In order to explore the scope of the self-assembly mechanisms of nucleic acids, we investigated aqueous solutions of DNA oligomers of length $N_B \leq 30bp$, into which various modes of random sequencing have been introduced, such that at selected positions in the sequence, the 4 primary A, C, G, T nucleobases are found with equal probability. Example of such "ranDNA" oligomers are indicated

in Figure 4.1, where we denote a randomly chosen base with the character "N" in the sequence. For example, 8N indicates the set of 8mers having 8 randomly chosen bases and C(8N)G the set of 10mers having complementary C-G terminal base pairs and 8 internal randomly chosen bases. In both cases the family of molecules corresponds to $4^8 \approx 65000$ different sequences mixed together in the same solution. The collective state of any such system crucially depends on the role of the polydispersity of interaction originating from the wide range of pairing strengths and pairing motifs. Well-matched oligomers may promote liquid crystal phases, contrasted by oligomers forming duplexes with low end-to-end interactions. On the other hand, there is a significant possibility of chains acting as crosslinks, the effect of which would be to take the system to a random isotropic gelated state similar to hydrogels formed by long DNA. It was thus difficult to anticipate the solution behavior of ranDNA.

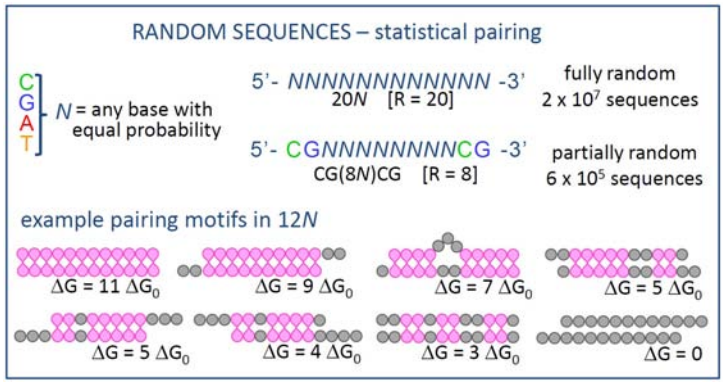
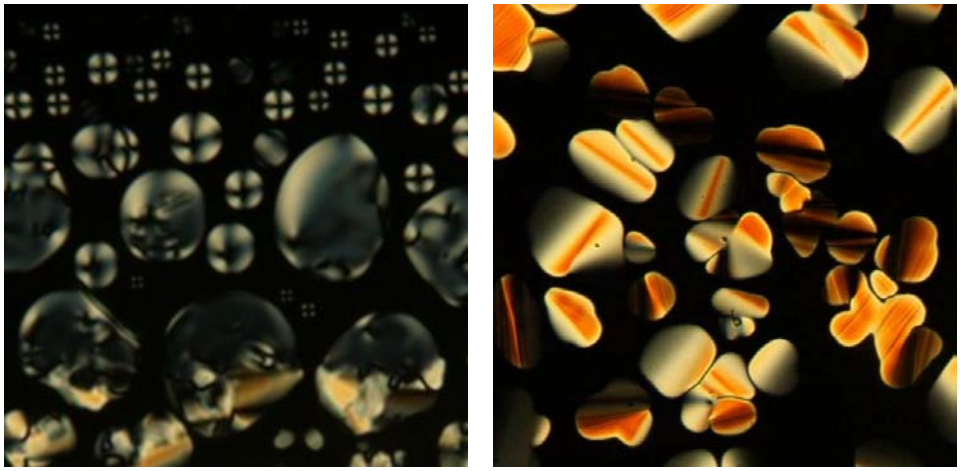


Figure 4.1: Random DNA (ranDNA) sequences are synthesized by randomly choosing one of the four basic nucleobases at given positions along the chains, and are thus mixtures of 4^R different sequences, where R is the number of randomly chosen bases. Because of the large number of sequences, ranDNA forms duplexes with a variety of distinct pairing motifs, the distribution of which is controlled by their binding energy ΔG, given in units of ΔG_0, the mean binding energy of a quartet of complementary bases, and by their lifetime $\tau \sim exp(\Delta G/\Delta G_0)$

Sequence description, ranDNA synthesis, HPLC purification and MALDI characterization can be found in the Chapter 3. In each case solutions were prepared in pure water with c_{DNA} in the range $300 < c_{DNA} < 1000 mg/ml$, enclosed in cells made of glass slides spaced by $10\mu m$, and examined in Depolarized Transmitted Light Microscopy (DTLM). Concentrations were determined as described in 3. LC ordering is heralded by the appearance of birefringent domains of characteristic color and texture that enable identification of the phases [1] (Figure 4.2).

4.2 LC Ordering of oligomers with designed sequence errors

We have first investigated the phase behavior of oligomers with designed sequence errors in the base pairing (Figure 4.3). The systematic study of the effect of such errors on the LC phase formation indicates that even a single terminal mispairing reduces the



(a) (b)

Figure 4.2: A variety of ranDNA solutions self-assemble into LC phases, recognized by the textures in thin cells observed in depolarized transmission optical microcopy (DTLM). (a) Chiral Nematic phase (N*) of 5′-12SCNNN-3′; (b) Columnar phase of 5′-20N-3′.

1	5'- CCTCAAAACTCC -3' 3'- GGAGTTTTGAGG -5'		complementary	N* and COL
2	5'- CCTCAGGACTCC -3' 3'- GGAGTCCTGAGG -5'		internal mismatch	N* and COL
3	5'- CCTCGGGGCTCC -3' 3'- GGAGCCCCGAGG -5'		internal mismatch	COL only
4	5'- ACTCAAAACTCC -3' 3'- GGAGTTTTGAGG -5'		terminal mismatch	COL only
5	5'- AATCAAAACTCC -3' 3'- GGAGTTTTGAGG -5'		terminal mismatch	COL only
6	5'- ACTCAAAACTCA -3' 3'- GGAGTTTTGAGG -5'		terminal mismatch	NO LC phases
7	5'- TCCTCAAAACTC -3' 3'- GGAGTTTTGAGG -5'		shift / overhang	NO LC phases

Figure 4.3: Liquid crystal phases of sequences with designed pairing errors. Columns indicate: sequential numbering of the sequences, sequences used in the experiments, schematics of their pairing motif, description of their pairing, LC phases observed in these solutions by exploring various concentrations at $T = 25^{\circ}C$.

self-association of the duplexes, as it might be expected. Internal errors are less problematic, affecting the phase behavior only when there are more than two in a duplex. In general, the N* phase is suppressed first, followed by the COL phase, as the number of pairing errors is increased. In solutions of sequences paired in a shifted mode with overhangs, LC ordering is stabilized if the overhangs are mutually complementary [44], but suppressed if the overhangs are non-complementary.

4.3 LC ordering of random sequence DNA

Figure 4.4a summarizes the phase behavior at room temperature of ranDNA as a function of the oligomer length N_B and of the number of random bases (R) when randomness is progressively increased from the center of the sequences. $R = 0$ corresponds to fully self-complementary (SC) sequences. We find that:

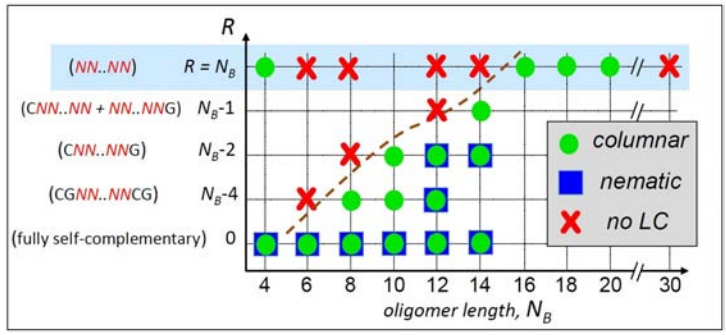
- the shortest sequence for which we find LC ordering increases with R, as indicated by the dashed line;
- the N* phase is more easily suppressed by randomness than the COL phase;
- quite remarkably, COL ordering can be found in fully random sequences (light blue shading), but only for $N_B \geq 16$, also confirming that the destabilizing effects of randomness are somehow mitigated by increase of length.
- This trend is limited, however, as LC ordering is not found for $N_B \geq 30$.
- LC ordering of random 4mers has been recently observed and we'll discuss it in Section 4.8.

The optical textures of the cells and local birefringence indicate that the N* and COL LCs of ranDNA are the same phases as those observed in complementary duplexes aggregating by end-to-end stacking.

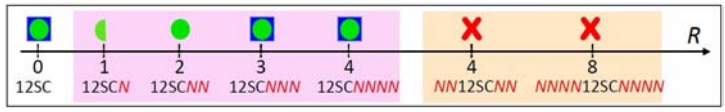
Figure 4.4b summarizes the phase behavior observed when random bases are added to the terminals of a complementary core, in this case the "Dickerson Dodecamer" self-complementary 12mer (12SC), a very well characterized LC-forming [80] B-DNA-type structure. Random sequences added to the 3' end only (pink shading) lead to 12SC duplexes with random overhangs (length l). In this case, while short random overhangs ($l = 1$ [–N] or $l = 2$ [–NN]) cause the loss of the N* phase (with the COL phase found in the case of $l = 1$ in a very narrow range of c_{DNA} and T), we interestingly observe the full LC behavior (COL and N*) for larger $l = 3$ and $l = 4$, a behavior closer to that of fully complementary overhangs than that of non-complementary overhangs, e. g. –T, –TT, which always suppress LC ordering [1]. We also explored the effect of random sequences of equal length added to both 12SC terminals (orange shading), a structure yielding duplexes with two random tails at each end. In this case, the LC ordering is disrupted for short and longer tails (–NN, –NNNN).

4.4 Phase behavior of random 20mers (20N)

The most surprising finding of the lot is the observation of LC ordering in solutions of 20N, the 20-mers having all bases randomly chosen. A 20N sample corresponds to a molecular population of about 1012 different sequences. Since we can assume that in 20N all sequences are equally represented, apart from statistical fluctuations (see Chapter 3), in a typical solution used in the experiments ($c_{DNA} \sim 750mg/ml$), a given sequence has a molar concentration of the order of $0.1pM$, its fully complementary potential partners being at a mean mutual separation of about $10\mu m$. The LC phase behavior



(a)



(b)

Figure 4.4: Maps of the LC phases observed in ranDNA of different total number of bases N_B and of random bases, R, either internally (a) or at the terminals (b). Symbols indicate the presence of the nematic (blue square) and of the columnar (green circle) phases in some range of concentration and temperature of the solution. The half green circle indicates that the columnar phase is present in a very narrow range of conditions (low T, large c_{DNA}). Shaded regions highlight specific phase behavior: light blue - fully random sequences; pink - double helices with a single overhanging random tail per duplex terminal; and orange - two random tails per duplex terminal.

of the 20N was studied by measuring the fraction of the cell volume filled by LC domains, ϕ_{LC}, vs. c_{DNA}, following heating to the ISO phase and cooling back to $T \sim 25°C$. The ISO phase fills the cells for small c_{DNA}, while the COL domains appear and coexist with the ISO for $c_{DNA} > 550mg/ml$. That is, the COL domains never grow to fill the whole area, with ϕ_{LC} increasing up to $\phi_{LC} \sim 0.7$ for $c_{DNA} \sim 1000mg/ml$, as shown by the open red circles in Figure 4.5.

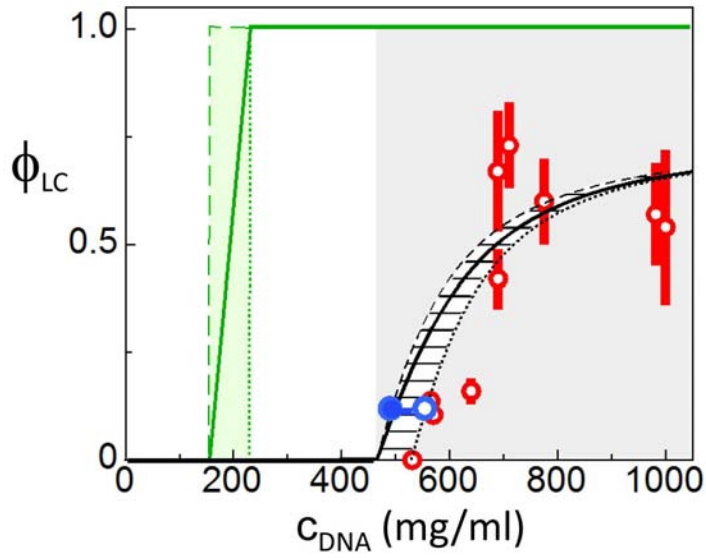


Figure 4.5: Experimental characterization of the self-assembly and LC formation of 20N. Red dots - LC-ISO phase coexistence at $T = 25°C$ as determined by the cell area fraction filled by the LC phase, ϕ_{LC}, as a function of the 20N concentration c_{DNA}. green line and shading - LC-ISO phase coexistence of a self complementary 20mer. full and empty blue dots - respectively the concentration of coexisting ISO and COL phases as measured by UV absorption on a capillary as in Figure 4.9, where macroscopic phase separation was forced through centrifugation. grey shading - LC-ISO phase coexistence range of 20N, black dashed and dotted lines - concentration of the coexisting ISO and COL phases.

This phase coexistence was further explored by using centrifugation to force macroscopic ISO/COL phase separation of 20N in capillaries (see Figure 4.9). COL domains are in this way compacted in the bottom of the capillary, with the ISO phase floating on top, enabling measurement of ϕ_{LC}. Capillaries were then cut so as to extract a known volume of each phase, which was then diluted to enable measurement of c_{DNA} via UV absorption (see Chapter 3). The values of the DNA concentrations for coexisting COL and ISO found near the threshold of phase coexistence ($c_{DNA} = 520mg/ml$, $\phi_{LC} = 0.17$) are plotted in Figure 4.5 as open and full blue squares, respectively. Their concentration difference, $\delta c_{DNA} \sim 80mg/ml$, is much smaller than the phase coexistence range in c_{DNA}, a typical feature of the phase diagram of multicomponent or polydisperse systems. This is in contrast to the behavior of solutions of single component oligomers, such as the complementary 20SC, where the coexistence range is the same as the concentration difference between the two phases and ϕ_{LC} approaches 1 (see green lines in Figure

4.5). The (red) vertical bars in Figure 4.5 indicate the variation of ϕ_{LC} for different thermal histories and cooling rates. This dependence is rather weak, with similar ϕ_{LC} found for both rapid and slow cooling of isotropic solutions. Since rapid cooling is too fast for complementary partners capable of perfect duplexing to find one another, imperfectly paired oligomers must be the basis of LC ordering in solutions of random 20mers.

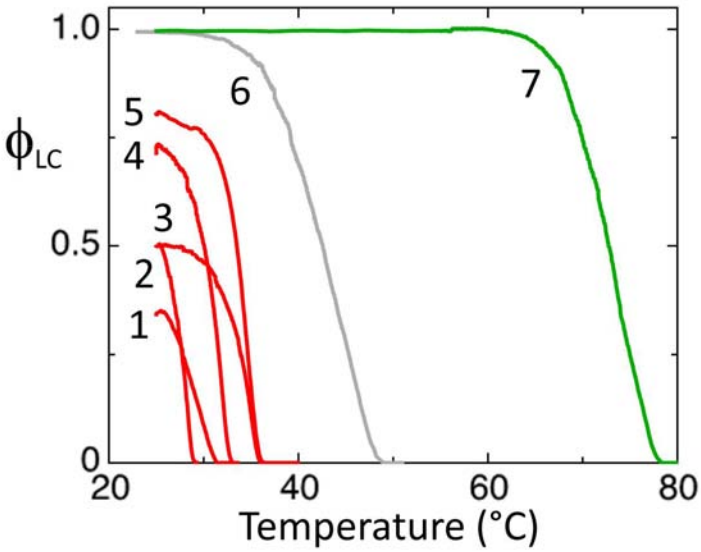


Figure 4.6: LC volume fraction ϕ_{LC} vs. temperature T. red curves (1-5) - progressive melting of the 20N COL phase in cells of various ϕ_{LC} in Figure 4.5. Curves 2 and 3 are obtained with the same cell, curve 3 with longer thermalization time at room T. grey curve (6) - COL-ISO ϕ_{LC} vs. T for 12SC at $\sim 1200mg/ml$. green curve (7) - COL-ISO ϕ_{LC} vs. T for 20SC at $\sim 600mg/ml$.

The LC domains melt as T is raised. Examples of the melting curve ϕ_{COL} vs. T are shown in Figure 4.6, where they are compared to the melting of the COL phase of 12SC (grey line) and of a self-complementary 20SC (green line). Various melting curves for 20N are shown as red lines the figure, corresponding to various concentrations and thermal histories of the samples. Quite evidently, the COL melting T (T_{COL}) of the 20N takes place at a much lower T than the 12SC, a clear indication that 20N interduplex interactions are weaker than those acting between well-paired blunt-ended duplexes.

The relatively low thermal melting temperature of the COL phase of 20N implies weaker duplexing and/or aggregation than in the fully complementary 20SC. The progressive disruption of duplexes can be monitored by measuring the intensity, I_F, of the fluorescent emission of Ethidium Bromide (EtBr) at low concentration in the 20N solution. In Figure 4.7 we plot I_F measured in COL and isotropic (ISO) phases in a cell where the two phases coexist. Above T_{COL}, marked by a dashed line, I_F decreases smoothly indicating that the 20N helix unbinds at a temperature $T_U \sim 55°C$. At $T < T_{COL}$, because of the anisotropy of EtBr fluorescence, I_F measured in the COL phase depends on the LC orientation. However, by appropriately averaging, we can extract a value of $\langle I_F \rangle (COL)$ (red line) that, in absence of other factors, should equal $I_F(ISO)$. We instead

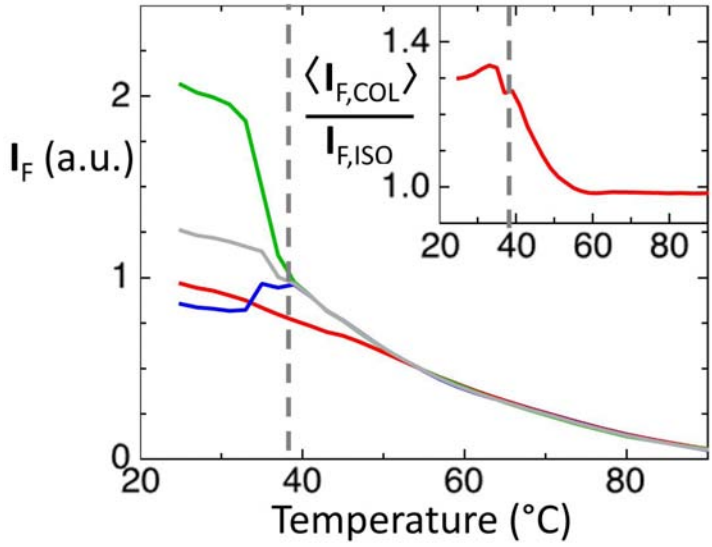


Figure 4.7: Fluorescent emission I_F of EtBr in a 20N solution ($c_{DNA} \approx 700mg/ml$) vs. T as a probe of duplex unbinding, measured in both the coexisting COL (blue, green lines) and ISO (red line) phases. Because of the polarized fluorescent emission of EtBr, I_F depends on the orientation of the COL ordering, parallel (blue line) and perpendicular (green line) to the cell plane. The mean, orientation independent, fluorescent emission in the COL phase ($\langle I_{F,COL} \rangle$, grey line) can be obtained from the weighted average of the parallel and perpendicular I_F. The inset shows the excess of mean fluorescent emission of the COL phase with respect to the ISO phase.

find $\langle I_F \rangle > I_F(ISO)$ (see inset in Figure 4.7). Since data in Fig. 4.7 are taken with increasing T, the descrease of $\langle I_F \rangle / I_F(ISO)$ as T grows above $T = 40°C$ reflects both the loss of the better quality of pairing promoted within the COL phase, which is lost as the LC melts, and the larger DNA concentration of the COL phase, slowly diffusing out.

Figure 4.8 uses I_F to compare the duplex melting of 20N, measured in the ISO phase (red line 1), with that of 12N (orange dashed line) and of various SC sequences at similar concentration. Quite clearly the melting of 20N and of 12N are very similar and distributed over a larger T interval than complementary sequences and take place in a T range between the 10SC and the 12SC, much lower than that of the 20SC. This indicates that the pairing of 20N is far from complete, involving a wide distribution of energies. It also suggests an average degree of pairing in the order of 10 couples of adjacent base pairs for both 20N and 12N.

4.5 Duplexes with random tails

The observations reported here confirm that LC ordering of oligomeric DNA is mediated by linear aggregation. As errors and randomness are introduced at the duplex terminals, LC ordering is lost. The larger sensitivity of the N* ordering to duplex mispairings has been recently predicted to be an effect of reduced interparticle bonding energy in a

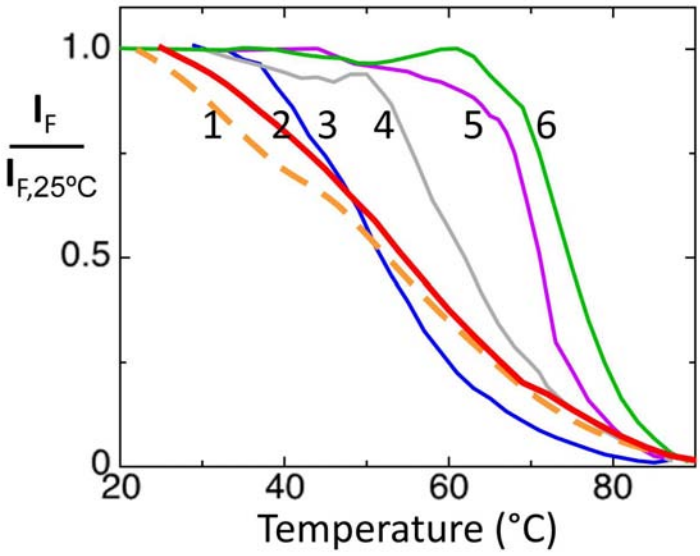
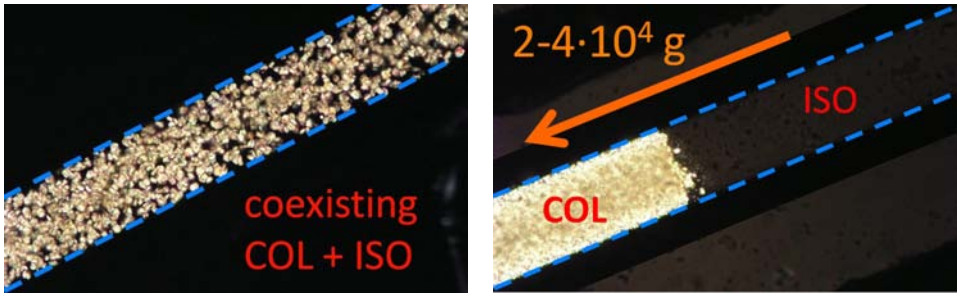


Figure 4.8: Number of duplexes remaining at T relative to the number at $T = 25°C$, extracted from I_F: orange dashed line (1) – 12N. red line (2) - 20N. blue line (3) - 10SC. grey line (4) - 12SC, purple line (5) - 16SC. green line (6) - 20SC. For all the sequences, the curves were obtained in the concentration range $600 - 800mg/ml$.

model of linearly aggregating cylinders [35], and it is also expected by theories describing the phase behavior of long rods as their flexibility increases [30, 81]. Moreover the higher packing density of COL leads to an increase in the lifetime of non-complementary duplexes, thus reducing the disordering effects of randomness.

It is therefore not surprising that duplexes with a stable core and four random tails (one per sequence terminal, orange set in Figure 4.4b) do not develop LC ordering, since in these cases the probability of attaining terminal pairing is small. More interesting is the behavior of duplexes with random overhangs (pink set in Figure 4.4b). The addition of a $l = 1$ random overhang strongly destabilizes the LC ordering. This is expected since for a $l = 1$ tail only $1/4$ of the casual collisions with other duplex terminals leads to formation of paired bases and thus to interduplex attractive interaction. As l increases, however, the probability that overhangs mutually bond because of random collision also increases. The average interduplex bond free energy $\langle \Delta G_{ID} \rangle$ can be estimated by computer generating collisions between random overhangs. ΔG_{ID} for each pair of colliding overhangs is evaluated on the basis of a simplified version of the standard nearest-neighbor model for DNA hybridization free energy [12] in which we attribute a free energy ΔG_0 to each adjacent pair of complementary bases (quartet), while non-complementary bases do not contribute to the duplex binding free energy (see Chapter 3). Here we assume $\Delta G_0 = 2.7 k_B T$, the room temperature DNA binding energy per base quartet averaged over all properly paired nucleobases. The resulting mean inter-duplex bonding free energy ranges from $\langle \Delta G_{ID} \rangle \approx 1.35 k_B T$ for $l = 1$, to $\langle \Delta G_{ID} \rangle \approx 2.1 k_B T$ for $l = 4$ (Figure 4.13, right axis). These quantities should be compared with $5 k_B T$ of



(a) (b)

Figure 4.9: DTLM image of 0.5 mm diameter capillary of 20N in solution, showing COL domains: (a) before and (b) after centrifuging at 16800 to 37800 rcf for as little as 30 minutes at $T = 22°$.

blunt end stacking and $6k_BT$ for interactions via mutually complementary overhangs of length 2 [44].

4.6 Non-ergodic association

With this background we can discuss the behavior of fully random sequences, and of the 20N in particular. When in solution, random sequences produce a variety of paired couples. An approximate calculation of the equilibrium distribution of all possible pair associations within the pools of ranDNA oligomers, can be obtained as a Boltzmann distribution:

$$P(\Delta G) = [n(\Delta G)] \exp(-\frac{\Delta G}{k_B T}) \tag{4.1}$$

where ΔG is the intraduplex binding energy, evaluated with the approximation that $\Delta G = \Delta G_0$ for every WC paired quartet and $\Delta G = 0$ for mismatched quartets (see Chapter 3 for more details). $n(\Delta G)$ is the number of possible oligomer combination yielding a given value of ΔG. $n(\Delta G)$ results from considering all possible pairs of sequences combined with all possible shifts l. This enumeration encompasses various pairing motifs, some of which are sketched in Figure 4.1. The resulting $n(\Delta G)$, is shown in Figure 4.10 for 6N (green dots), 12N (grey diamonds) and 20N (blue squares). The spread in energy of the resulting equilibrium distributions (Figure 4.11) at $T = 25°C$, confirms the expected variety of pairing motifs in ranDNA. The 20N mean duplex binding free energy calculated from the equilibrium distribution is $\langle \Delta G \rangle_{EQ} \approx 17\Delta G_0$, a value nearly as large as $19\Delta G_0$, the maximum binding energy for fully-paired 20mers. However, the 20N duplex melting behavior noted above (Figure 4.8) suggests instead a mean intraduplex binding energy to be of the order of $\Delta G \approx 10\Delta G_0$. This indicates that equilibrium distributions are generally not adequate to describe the association of ranDNA.

In trying to understand this discrepancy we realized that kinetics of pairing plays a crucial role that must be considered in modeling the pairing distribution. Indeed, the lifetime τ of well-paired DNA oligomers can easily exceed the typical times involved in

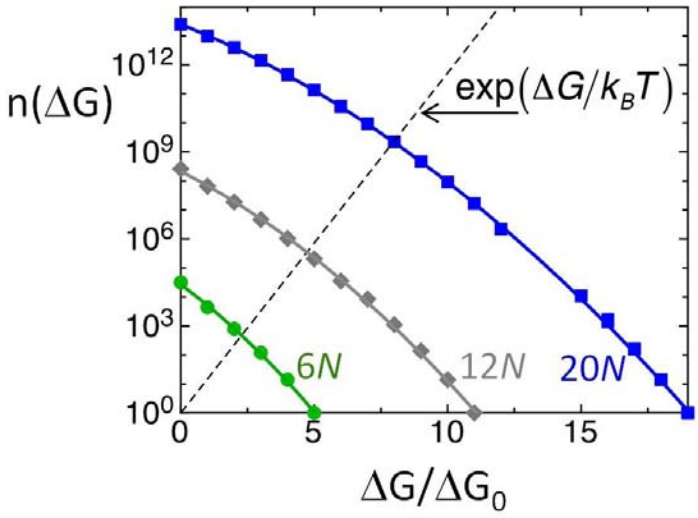


Figure 4.10: Number $n(\Delta G)$ of duplexes differing in sequences or in shift that can be formed within the ensemble of fully random sequences of a given length as a function of the pairing free energy ΔG. blue squares - 20N; grey diamonds - 12N; green dots - 6N. The $n(\Delta G)$ are normalized to $n(\Delta G) = 1$ for the largest energy for each given oligomer length. black dashed line - ΔG dependence of the Boltzmann factor, on the same scale. Free energy is expressed in units of ΔG_0.

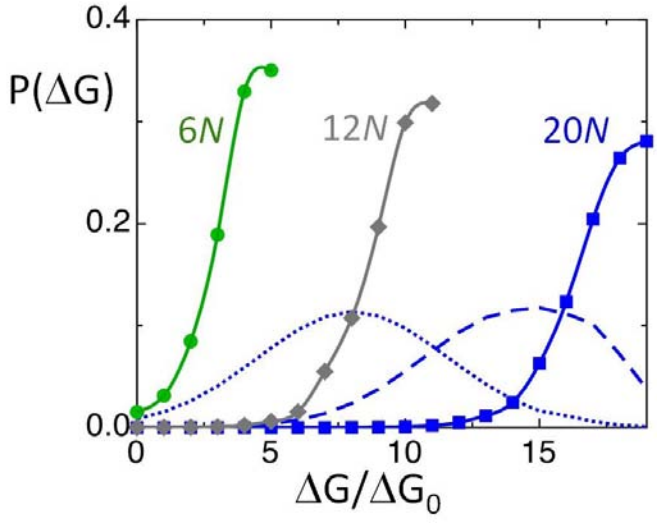


Figure 4.11: Equilibrium distribution $P(\Delta G)$ of the intraduplex binding free energy in fully random ranDNA at $T = 25^{\circ}C$. blue squares - 20N. grey diamonds - 12N. green dots - 6N. dashed and dotted blue lines - $P(\Delta G)$ for 20N at $T = 45^{\circ}C$ and $T = 60^{\circ}C$, respectively.

experiments. At room temperature, complementary 8mers have lifetimes $\tau \sim 1s$ [82], while complementary 20mers have lifetimes $\tau > 10^5 s$ [83**?**]. If we assume an activated

behavior to express these complementary duplex lifetimes as a function of free energy:

$$\tau(\Delta G(N_B)) = \tau_0 \exp(\frac{\Delta G(N_B)}{k_B T}) \tag{4.2}$$

where $\Delta G(N_B) \approx (N_B - 1)\Delta G_0$ is the duplex binding free energy of well-paired duplex strands, we find a mean attempt time $\tau_0 \sim 10ns$, enabling us to estimate the lifetimes of duplexes vs. their binding free energy ΔG.

The strong dependence of τ on ΔG is the critical factor in determining the thermodynamic status of ranDNA solutions. Oligomers collide and interact, the actual energy level attained in each interacting pair being determined by the number and location of well-paired bases. Collisions yielding weak binding originate short-lived pairs that rapidly separate and proceed to further collisions. Therefore, short sequences ($N_B \lesssim 12$), whose binding and unbinding take place on time scales shorter than the experimental time, can reach equilibrium. As longer ($N_B \gtrsim 12$) sequences are considered, the situation changes since they can form duplexes having a lifetime comparable or larger than the experimental time, which are therefore effectively stable. This fact prevent the system from exploring in time all the possible states, in this way the equilibrium distribution is never reached (non-ergodic behavior).

We modeled the kinetic behavior of ranDNA by computer generating successive encounters between one given N-mer sequence and randomly generated N-mers, evaluating ΔG and the lifetime $\tau(\Delta G)$ for each event, and then summing these lifetimes until a given total time τ_{TOT} was reached (see Chapter 3). The energy distribution of the ensemble of duplexes found at a time $\tau_{TOT} = 10hr$ is shown in Figure 4.12. The precise choice of τ_{TOT} is not very critical because of the strong dependence of τ on ΔG: a choice of τ_{TOT} within the range $1hr < \tau_{TOT} < 100hr$ displaces the distribution of at most $\pm 1\Delta G_0$.

Upon comparing the equilibrium and kinetically arrested distributions (see Figure 4.12), it appears that the 6N and the 12N are at equilibrium, or very close to it, while the 20N kinetic distribution is markedly different from equilibrium and it is characterized by a mean energy value $\langle \Delta G_{KIN} \rangle \approx 11\Delta G_0$. These results are in good agreement with duplex melting observations (Figure 4.8), and indicate that, in the case of 20N, the approach to equilibrium and to stronger binding is impaired by kinetic arrest.

The non-equilibrium behavior observed in the simulations opens a path to understanding the formation of LC phases of 20N. In Figure 4.13 we show the distribution in overhang length obtained from the equilibrium (blue dots) and non-equilibrium (red open squares) duplex distribution of 20N. As clearly visible in the Figure, the equilibrium distribution favors well-paired duplexes and yields an average overhang length $\langle l \rangle \approx 0.5$ whereas the non-equilibrium ensemble of kinetically trapped duplexes is characterized by a broad distribution of overhang length with $\langle l \rangle \approx 2.5$. Therefore, successive random 20N collisions lead to a kinetic arrested population of duplexes having a central reasonably well-paired stretch with terminating stretch of overhanging random bases, as sketched in Figure 4.14. Their condensation into LC domains is thus readily understood by the similarity of such a system to the 12SC with random overhangs (Figure 4.4, pink set) discussed above. Indeed, the low melting T for the COL phase of 20N

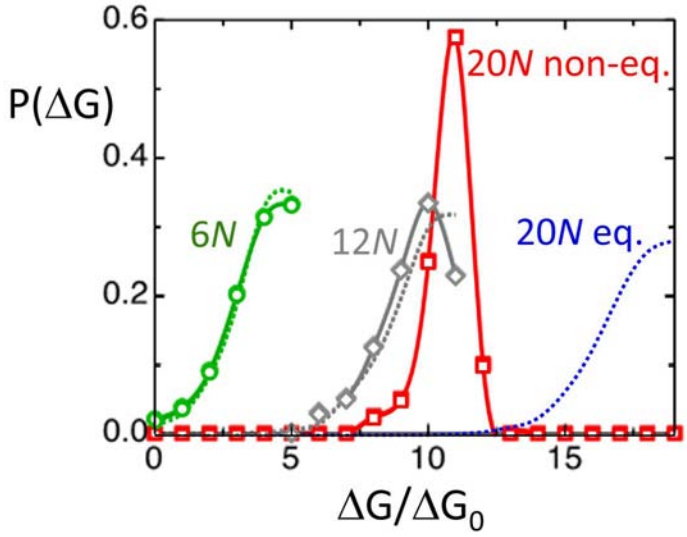


Figure 4.12: Free energy distribution $P(\Delta G)$ calculated through kinetic evolution on the basis of duplex lifetime and random collisions. red squares - 20N. grey diamonds - 12N. green dots - 6N. Dotted lines repeat, for comparison, the equilibrium distributions in (b). While 6N and 12N are at equilibrium or nearly so, the distribution of 20N is kinetically arrested and far from equilibrium.

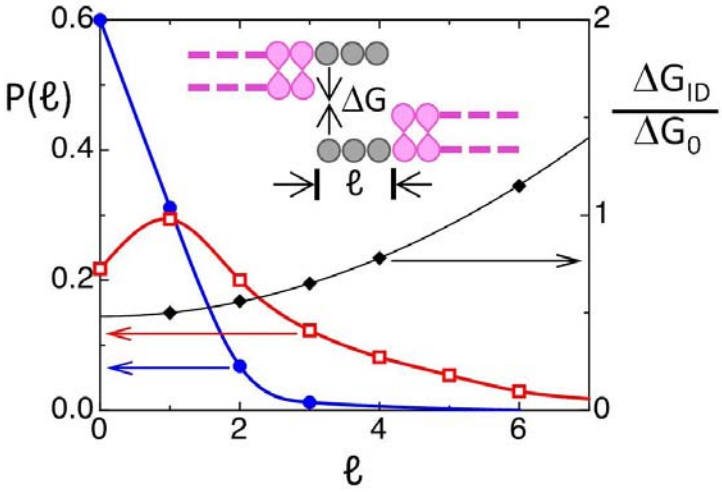


Figure 4.13: Left axis: Calculated overhang length distributions $P(l)$ for 20N. blue dots - equilibrium distribution. red open squares - kinetically arrested distribution, on the same scale. Right axis: black diamonds - mean interduplex interaction free energy, calculated as the average value of the binding free energy resulting from collisions of random overhangs.

further confirms this picture of duplexes interacting via the weak and, importantly, annealable end-to-end coupling provided by random tail interaction. From the evaluations above we expect such coupling to be of the order of $1.6k_BT$, thus significantly weaker

than blunt-end interactions. Moreover, the strengthening of interduplex interactions observed upon aging the samples (see Figure 4.6) supports the notion that the interaction between adjacent duplexes can be improved by enhancing the matching of contacting overhangs, a process expected to take place by duplex flipping and hopping within the COL structure.

The same analysis performed on 12N reveals the importance of the oligomer length in the formation of LC phases. Equilibrium and kinetic distributions are in this case quite similar, both yielding $\langle l \rangle < 1$, a picture confirmed by the fact that T_U for 12SC and for 12N are not very different (Figure 4.8). This observation, when combined with the significant probability of finding pairing errors at the duplex ends, justifies the fact that end-to-end interactions in 12N are too weak to support linear aggregation and LC phase formation.

The formation of duplexes in 20N can thus be regarded as a paradigmatic example of a system diffusing in a space populated by energy traps that are deep enough to produce non-ergodic behavior [84, 85]. This is a conceptual frame intensively studied to account for non-ergodicity and aging in condensed matter systems of various kinds, including disordered systems and polymers with multiple folded states. A remarkable consequence of the combination of factors at play, sketched in Figure 4.14, is that ranDNA develops LC long range ordering only in a limited interval of lengths near $N_B \sim 20$ (Figure 4.4). In the 20N, kinetic arrest produces duplex pairs with interpair interactions that are sufficiently weak to enable annealing and equilibration into LC domains. Shorter ranDNA oligomers form a population of equilibrium duplexes that lack the interactions necessary for self-assembly and LC formation. At the opposite extreme, when the sequences are longer, the random overhanging tails can form additional kinetically arrested interduplex interactions, leading to oligomer networking and gelation [86] and suppressing LC formation. Observations show that ranDNA with $N_B \gtrsim 30$ yields viscous isotropic solutions without LC domains. Shearing between the cell plates with flow velocity at 45° to the DTLM polarizer gives transient optical transmission, indicative of transient birefringence, evidence for the formation of an isotropic gelated state.

It is of interest to question whether the nonergodic behavior of 20N can be avoided by thermal annealing at a temperature very close to the unbinding temperature T_U, as is the case in the known efficient annealing in Polymer Chain Reaction (PCR) systems that enables $N_B = 20$ primers to find their complementary target sequence on very long DNA. As it turns out, in the ranDNA case, because of the enormously larger number of nonideal pairing motifs, annealing closer to equilibrium generates a broad distribution of weakly bound pairs that are nonideal for LC formation. This is shown in Figure 4.11, where we plot the equilibrium distribution at $T = 45^\circ C$ (blue dashed line) and for $T = 60^\circ C$ (blue dotted line). Inspection of these curves clearly indicates that upon lowering the pairing and stacking energy (i.e. upon increasing T), the behavior 20N system is even more strongly affected by its huge variety of conformations expressed in $n(\Delta G)$.

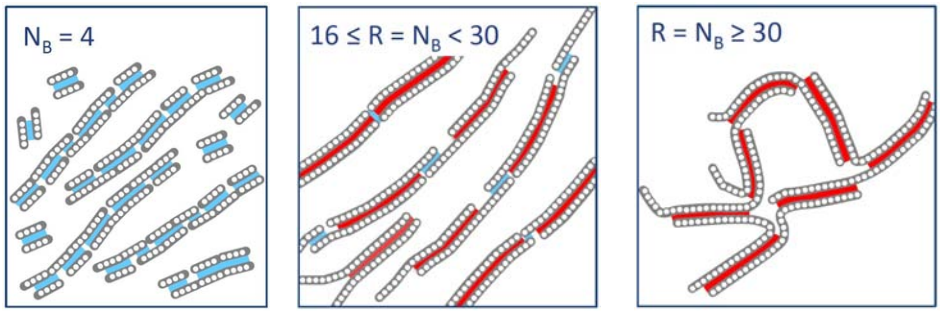


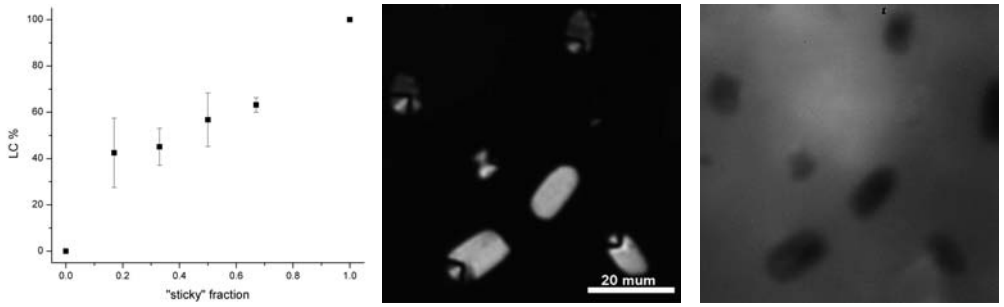
Figure 4.14: Sketches of the behavior of ranDNA oligomers of different lengths in solution. Blue shading: $\Delta G < 15\Delta G_0$; τ short; equilibrium binding. Red shading: $\Delta G > 15\Delta G_0$, τ long compared to experimental time; kinetically arrested binding. Thus, short oligomers (a) form an equilibrated ensemble of duplexes; oligomers of intermediate length (b) form kinetically arrested duplex cores with weakly mutually attractive tails enabling linear aggregation and LC ordering; long oligomers (d) form multiple kinetically arrested cross-linking bonds leading to gelation.

4.7 Polydispersity of attraction

The 20N COL-ISO phase coexistence range, in which the system never gets entirely into the COL phase at high c, is much broader than that found in complementary duplexes and also broader than the concentration difference δc_{DNA} of the coexisting phases, as shown in Figure 4.5. This is evidence that the polydispersity of 20N duplex structures affects the phase separation in a way comparable to that found in multicomponent lyotropics LCs, such as polydisperse mixtures of rods [87, 88]. In 20N, collisions and energy-dependent lifetimes give rise to a significant polydispersity of duplex structures, with larger and smaller overhangs (Figure 4.13), and a variable number of terminal mispairings. This yields a polydispersity of attraction that has the potential to induce phase separation, since the more strongly interacting duplexes give rise to longer and more stable aggregates, which in turn can more easily overcome the Onsager threshold [27].

To test whether the variety of end-to-end duplex interactions can indeed yield phase separation, we studied a mixture of 12SC-CG (i.e. 12SC with a self-complementary 2-base overhang at the 3′ end), providing definite end-to-end attraction, and of two mutually complementary sequences, one of which terminated at the 5′-end by a FITC fluorescent group that prevents stacking and pairing interactions at one duplex end. This mixture promptly phase separates into coexisting birefringent COL domains with low fluorophore concentration and ISO fluid with high fluorophore concentration, with ISO-COL coexistence over nearly the complete range of relative concentration (see Figure 4.15). This test highlights a new route for the self-assembly of DNA oligomers, previously unnoticed.

To the extent that this binary mixture of interacting and non-interacting duplexes can be considered a model for the polydispersity of attraction of 20N, the phase behavior described in Figure 4.5 can be interpreted as a manifestation of the intrinsic polydispersity



(a) (b) (c)

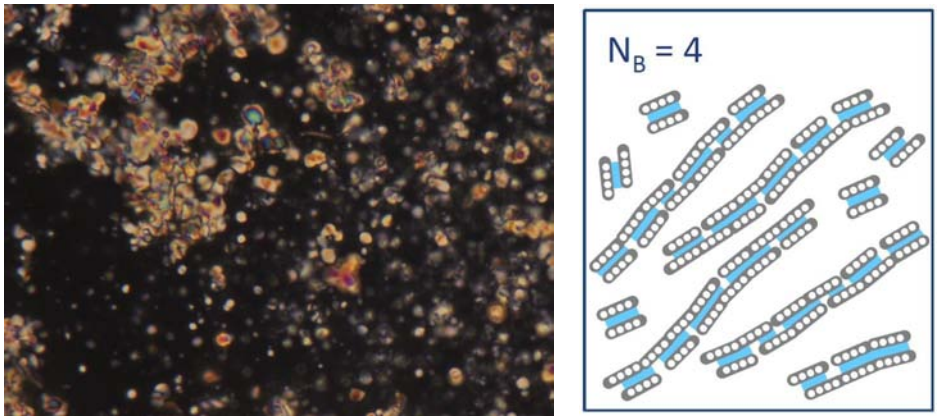
Figure 4.15: Phase behavior of mixtures of interacting and non-interacting duplexes: a population with sticky unpaired strands is mixed to helices with terminals impairing linear aggregation. (a) Average LC fraction as a function of the interacting/non-interacting ratio; (b) DTLM and (c) fluorescence images of a sample of 12SC-CG mixed with FITC terminated helices: the birefringent domains in (b) are depleted of the fluorescent, non-interacting helices.

of random-sequence DNA duplexes. Through this phase separation, the system self-selects chains capable, through long-lived intra-duplexes binding and annealed inter-duplex pairing, to organize in long physically-bound helices. This behavior unavoidably strengthens the notion that self-assembly of nucleic acids could have been instrumental in the formation of long double-helices on the basis of shorter random sequences.

4.8 The strange case of random 4mers (4N)

Recently we found evidences of LC ordering of random 4mers (4N). As shown in Figure 4.16a, at very low temperature, $T = 4°C$, and at $c_{DNA} \approx 600mg/ml$ COL textures in coexistence with ISO are found. This finding was a real surprise since this behavior cannot be explained with in the same way we interpreted the behavior of longer random oligomers with $16 \leq N_B < 30$. As we saw, longer random oligomers do not reach duplexing equilibrium because of the long lifetimes. However, as it is shown in Fig. 4.12, as the oligomer length is equal or less than 12, the distribution of duplex population is the equilibrium one. In this condition, however, the nature and distribution of duplex terminals does not provide enough end-to-end attraction to yield aggregates long enough to drive liquid crystal ordering. Hence, the ordering of the 4mers implies the existence of yet another regime of self-assembly. Very short oligomers, as 4mers, have a very low hybridization energy $\Delta G = -3.65kcal/mol$, that is larger but comparable to the energy involved in blunt or sticky end-to-end interactions. Thus the assembly of the oligomers cannot be described as a 2-step process as in the case of longer sequences, since the energy involved in the formation of double helices and in their interactions is of the same order. In this regime, DNA 4mers are able to combine into long chained structures leading to LC ordering, as discussed in Chapter 5 of this thesis.

Following this descriptions, 4N can be seen as a particular case in which the formation of long chained structures is affected by pairing errors due to the nature of the



(a) (b)

Figure 4.16: (a) DTLM image of COL phase in coexistence with ISO observed in random 4mer. Sample prepared at $C_{DNA} = 150mg/ml$ is left to dry in $50\mu m$ thick glass capillary at $T = 4°$. Estimated final concentration is $c_{DNA} \approx 600mg/ml$. (b) Sketches of the behavior of 4N oligomers: LC ordering is driven by the formation of chained structures with pairing errors.

random building blocks (see cartoon in Fig. 4.16b). A rude estimation of the stability of these structures can be obtained by comparing the thermal denaturation of random 4mers with that of 4mers with defined sequence (see section 5.3), as reported in Figure 4.17. The difference between $T_M \approx 60°$ of GCTA, forming chains of repeated sticky-end duplexes, and $T_M \approx 40°$ of 4N is understandable considering the lower degree of pairing of the latter.

Figure 4.18 shows the plot of the predicted T_M of 4mers upon progressively reducing the hybridization free energy by the presence of pairing mismatches. In the Figure 4.18, T_M decreases as the amount of free energy loss is increased. The value 1 in the figure indicates a loss of energy $\Delta\Delta G = \Delta G_0$, equivalent to one mismatched quartet per oligomer.

Additional contribution is taken into account to mimic the energetic cost for possible loss of coaxial stacking. Green curve is the predicted T_M of 4 bases duplex into a chained structure, i.e. considering the stacking contributions of the adjacent duplexes and assuming that those are not affected by pairing errors, red curve is the predicted T_M of 4 bases duplex alone, thus without stacking interactions with neighbouring duplexes. These two curves represent respectively an upper and lower bound for the blue curve, that is the predicted T_M of 4 bases duplex into a chained structure, in which pairing errors can affect both internal pairing and coaxial stacking. The required decrease of T_M from $60°C$ to $40°C$ is obtained for $\Delta\Delta G \approx 1.2\Delta G_0$. Since a pairing error of one base causes the missing of two quartets, and thus the decrease of the free energy of $\Delta\Delta G = 2\Delta G_0$, the finding of 1.2 correspond to one missing base pairing every two duplexes.

Despite the fact that we do not have yet developed an interpretation for the formation of LC in concentrated solutions of 4mers, the elements shown here strongly suggest that

the explanation of the phenomenon has to be searched in the fact that this ultra-short oligomers have a mode of association that involves at once duplexing and chaining.

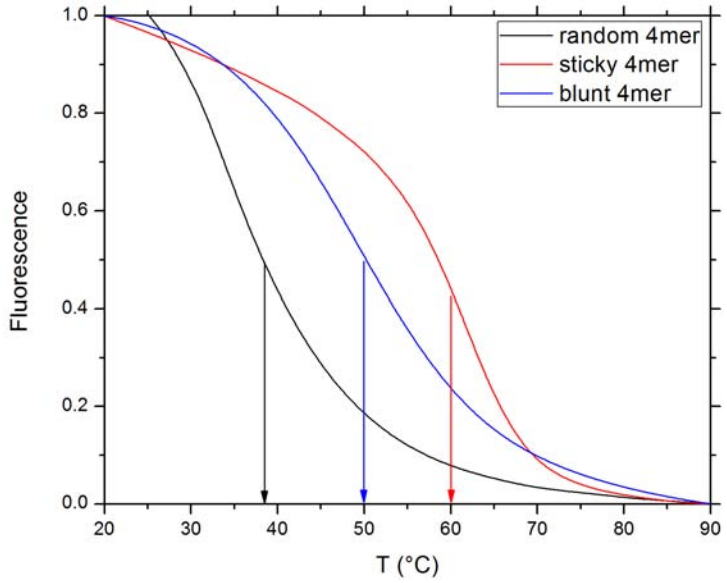


Figure 4.17: Comparison of experimental thermal denaturation curves of different 4 bases short DNA oligomers: black - random 4mer, red - sticky 5′-GCTA-3′, blue - blunt 5′-GTAC-3′. At high $c_{DNA} \approx 600 mg/ml$ formation of linear chained structures causes the high melting temperatures found for these sequences: $T_M \approx 40°C$ for 4N, $T_M \approx 50°C$ for 4blunt and $T_M \approx 60°C$ for 4sticky.

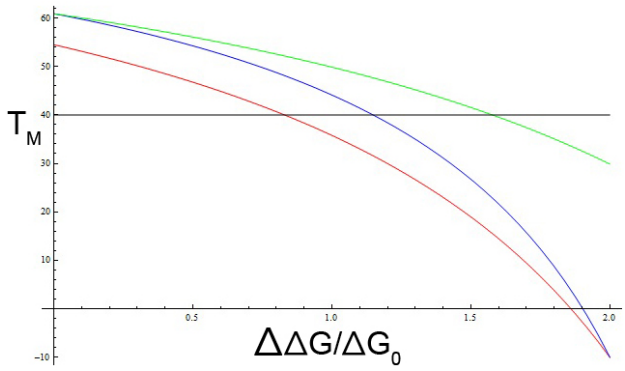


Figure 4.18: Plots of the predicted T_M of 4mers vs the variation of the binding free energy $\Delta\Delta G/\Delta G_0$ given by pairing mismatches. green - predicted T_M of 4 bases duplex into a chained structure; red - predicted T_M of 4 bases duplex alone; blue - predicted T_M of 4 bases duplex into a chained structure, in which pairing errors can affect both internal pairing and stacking. Interceptions with $T_M = 40°C$, black line, give the estimations of $\Delta\Delta G/\Delta G_0$ involved in the observed T_M variation.

4.9 Conclusions

The study of ranDNA has revealed that LC ordering of DNA oligomers in solution is found even in presence of a large amount of randomness. We find a range of lengths of random DNA sequences, between the isotropic fluid arrangement of short oligomers and the isotropic gel of long random DNA strands, where a rich combination of random pair formation, equilibrium annealing, kinetic arrest, phase demixing and mesophase ordering, yields a pathway toward long-range LC ordering.

Evidence indicates that solutions of oligomers with 20 randomly chosen bases can evolve into a population of kinetically arrested self-assembled pairs characterized by a structural theme that enables the formation of linear aggregates and promotes condensation into LC domains. Given the extreme – but at the same time controllable - heterogeneity of these systems, and given the remarkable combination of self-assembly processes that guide their behavior, we envisage ranDNA as a paradigm for the study of the effects of random interaction disorder on the collective behavior of self-associating molecules in solution.

LC ordering of 4 bases short ranDNA, instead, follow from the arrangement of these oligomers into long chained structures able to form LCs. The restoring of ranDNA self-organization into ordered structures even at low oligomers lengths, is a crucial finding in the context of the formation of nucleic acids in prebiotic scenarios.

CHAPTER 5

LC phases of 4 bases DNA oligomers

The ability of double stranded DNA to form liquid crystal (LC) phases when hydrated has been known since the late 1950s. Since that moment LC ordering of long nucleic acids, with number of base pairs $N_{bp} > 100$ has exhaustively been characterized. More recently it has been discovered that also shorter DNA oligomers, $6 \leq N_{bp} \leq 20$, show the tendency to order into LC phases in concentrated solutions. Even if these DNA strands are too short to fulfill the shape anisotropy requirements of the Onsager's criteria for LC ordering, in such system the key mechanism is the ability of short duplexes to self assemble through end-to-end adhesion into longer linear reversible aggregates, which allow the emergence of LC phases [1]. Here I report on the recent observation of LC phases in solutions of even shorter oligomers, i.e.4 bases DNA strands with different pairing motifs: 5'-GTAC-3', forming blunt-end duplexes, and 5'-GCTA-3', forming sticky-end duplexes.

There is a twofold motivation for exploring the formation of liquid crystal phases in even shorter oligos. On the one hand, understanding the limiting factor for the ordering of short oligomers is an important information in the context of the prebiotic implications of DNA self-assembly. The discovery of LC ordering in 4mers makes the onset of RNA world more plausible, as better discussed in chapter 6. On the other hand, the investigation of the liquid crystal ordering of ultra-short oligomers leads to the exploration of a new regime in the self-assembly of the nucleic acids with potential new physics and new phenomena, as it will be better discussed in this chapter.

5.1 Blunt-end GTAC LC phases

Self complementary sequence 5'-GTAC-3' hybridizes forming blunt-end duplexes that at high concentration and low temperature show a variety of liquid crystal mesophases. At $c < 400mg/ml$ the system is totally isotropic (ISO). Keeping fixed Temperature at $3°C$ and increasing concentration, at $c \approx 440mg/ml$, Cholesteric phase (N^*) starts to appear in coexistence with ISO (Figure 5.1), filling the whole volume in the whole T range $[0-5°C]$ at $c \approx 486mg/ml$ (Figure 5.2). Colored textures, typical of Cholesteric phase, are due to selective reflection from the superposed twisted nematic layers forming a macroscopic helical superstructure with pitch in the visible range. The color shift from blue to green at increasing concentration indicates that cholesteric pitch is lengthening, behaving against the closer packing of the nematic layers. At $c \approx 520mg/ml$ N^* - Columnar

Uniaxial (C_u) coexistence is found: characteristic focal conics grow between N^* textures (Figure 5.3). Cu phase extends to almost 100% of the volume at $c \approx 560 mg/ml$ (Figure 5.4) and it's replaced by higher ordered Columnar (C_2) at $c \approx 590 mg/ml$ (Figure 5.5). At very high concentration $c > 680 mg/ml$ only Spherulites and Soft Crystals (Cry) in coexistenxe with ISO are found (Figure 5.7). The peculiarity of this phase diagram is the finding of a very unusual mesophase in coexistence with C2, its textures are *mosaic* type with very low birefringence typical of the Blue phases (Figure 5.6).

Increasing the temperature the system melts into ISO phase with increasing T_m with larger Concentrations. N^* melts in the range $3°C < T_m < 4°C$, C_u at $5°C < T_m < 10°C$, C_2 in the range $12°C < T_m < 20°C$ and Crystalline phases are found even at high temperature $T_m > 25°C$. Blue phase has $5°C < T_m < 8°C$. Complete c - T phase diagram is reported in Figure 5.8.

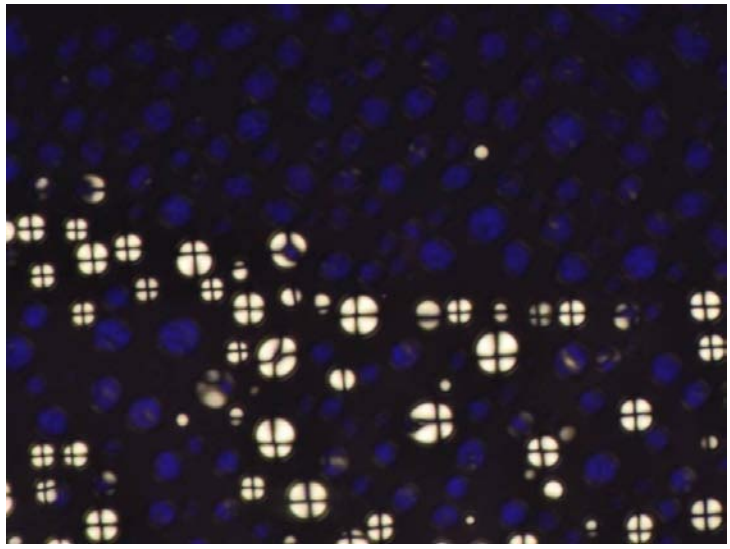


Figure 5.1: Cholesteric Phase (N^*) in coexistence with Isotropic of blunt-end 5′-GTAC-3′ at $c \approx 440 mg/ml$ and $T = 3°C$. Blue color is due to selective reflection.

5.2 Sticky-end GCTA phase diagram

Sequence 5′-GCTA-3′ can hybridize forming sticky-end type duplexes with a core of 2 paired bases and 2 bases sticky tails. The core can in principle be either GC or TA, respectively forming 2 families of non interacting -TA or -GC tailed duplexes. It is possible to calculate thermal stability of both double stranded structures via NN parameters [11, 12] (see Section 3.6). Contributions to Gibbs free energy come only from the core paired quartet GC/GC or TA/TA and two stabilizing dangling ends termination G/CT and CT/A. The total Gibbs free energy variation, between the unpaired and paired states, is given by the sum of these contribution and it turn out to be respectively $\Delta G_{GC} = -3.14 kcal/mol$ and $\Delta G_{AT} = -0.96 kcal/mol$. Furthermore, since the sticky

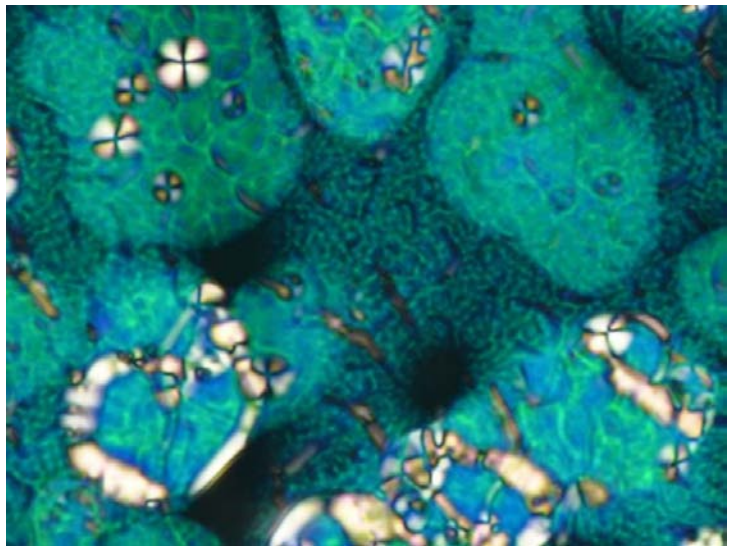


Figure 5.2: Cholesteric Phase (N^*) of blunt-end 5′-GTAC-3′ at $c \approx 490 mg/ml$ and $T = 3°C$. Colors are due to selective reflection in the blue-green region of the visible spectrum.

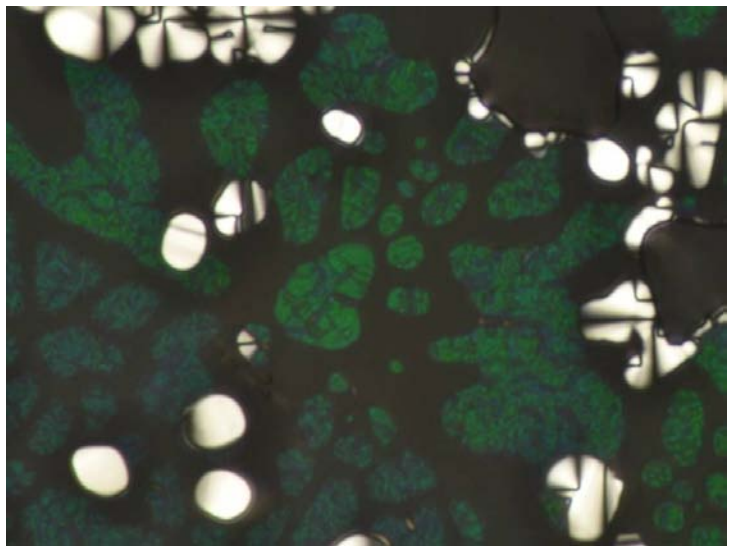


Figure 5.3: N^* - C_u coexistence of blunt-end 5′-GTAC-3′ at $c \approx 520 mg/ml$ and $T = 3°C$. Colors of the cholesteric domains are due to selective reflection in the green region of the visible spectrum.

tails have the same length and sequence of the core bases, ΔG of interaction of the sticky tails is exactly the same ΔG_{GC} and ΔG_{AT} of the cores. Thus the peculiarity of this system is the fact that there is no temperature difference between the duplexes formation and the tails interaction.

At concentration higher than $400 mg/ml$ and temperature below $15°C$ typical Liquid Crystalline phases are found: N^* from $c \approx 420 mg/ml$ (fig. 5.9) that is replaced from Col

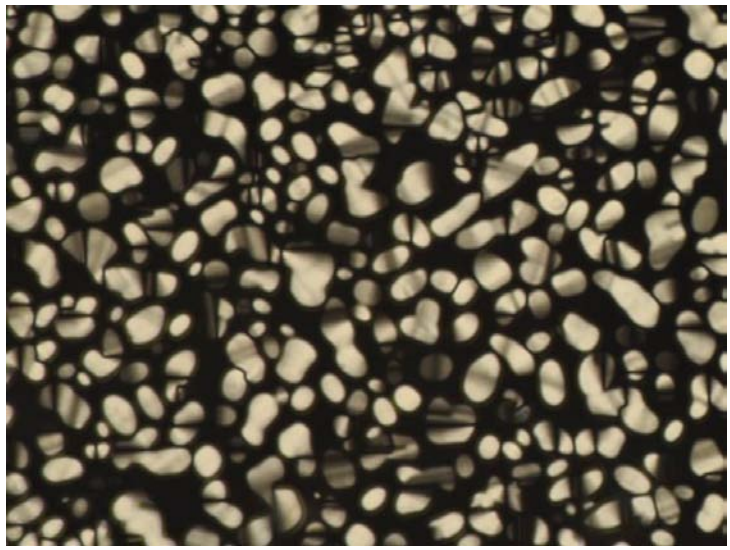


Figure 5.4: Columnar Uniaxial phase C_u of blunt-end 5′-GTAC-3′ at $c \approx 530 mg/ml$ and $T = 3°C$. Typical pseudo *focal conics* textures are clearly visible.

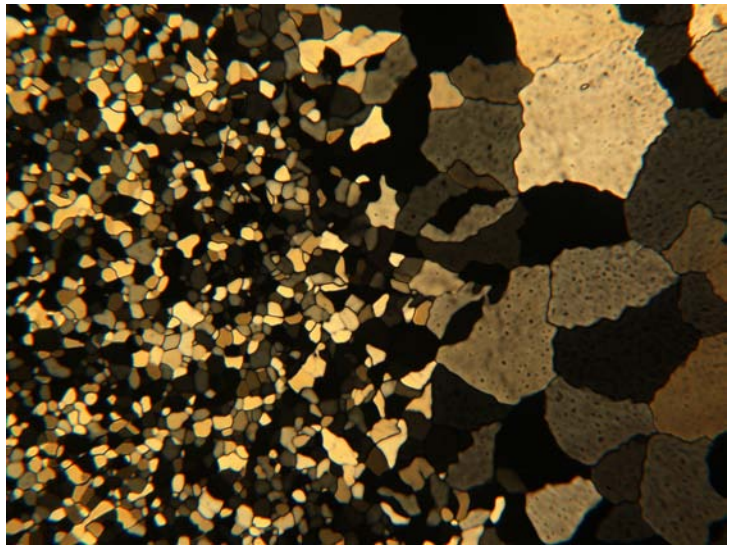


Figure 5.5: Higher-ordered Columnar phase C_2 of blunt-end 5′-GTAC-3′ at $c \approx 680 mg/ml$ and $T = 7°C$. Typical *mosaic* textures are clearly visible.

above $c \approx 540 mg/ml$. The only LC phase that resists even at room temperature is the highly ordered Columnar C_2: at $c \approx 750 mg/ml$ its melting temperature reaches $30°C$ (Fig. 5.13). It's still uncertain the definition of a C_u region because clear textures of this phase haven't been observed yet. Even if cooling from N^* formation of fan shaped *Focal Conics* has been found, it's not a clear proof of C_u phase (Fig. 5.10). Moreover $N^* - C_2$ boundary is usually found in concentration gradients without C_u formation in between

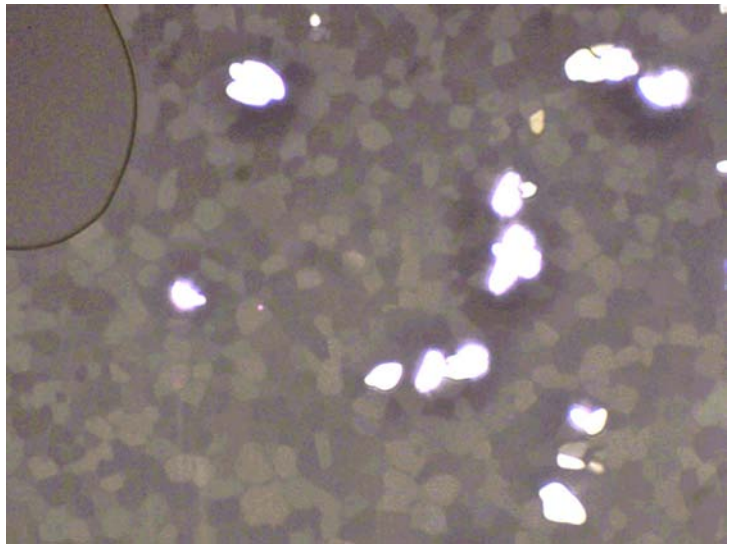


Figure 5.6: Bluish phase in coexistence with C_2 of blunt-end 5′-GTAC-3′ at $c \approx 590mg/ml$ and $T = 3°C$. Low birefringent *mosaic* textures are visible at high exposure time.

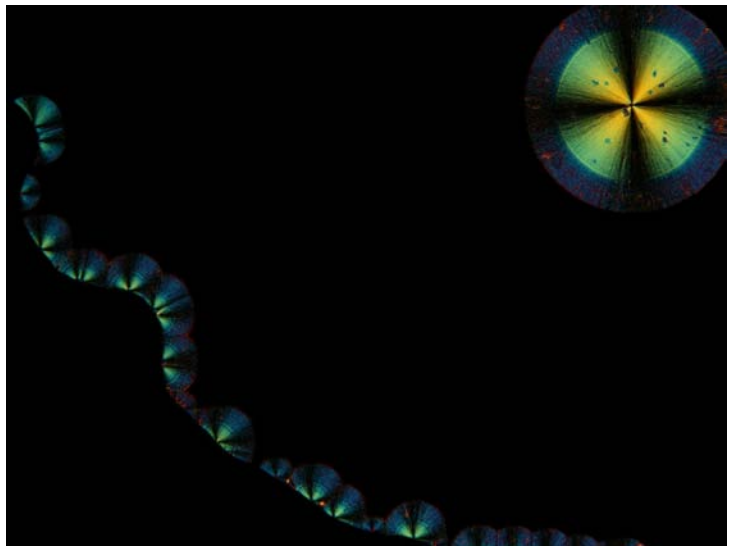


Figure 5.7: Spherulites nucleation from ISO of blunt-end 5′-GTAC-3′ at $c \approx 880mg/ml$ and $T = 28°C$. Typical crystallization occurs via the process of nucleation and growth, from fan-shaped seeds, of radially arranged, elongated domains with sharp grain boundaries.

(Figure 5.12).

Cholesteric phase instead shows all the typical textures previously observed with longer DNA oligomers, 6-20 base pairs [1]. In particular cholesteric pitch is found to decrease with increasing concentration and temperature (Figure 5.11), following the closer packing of the nematic layers.

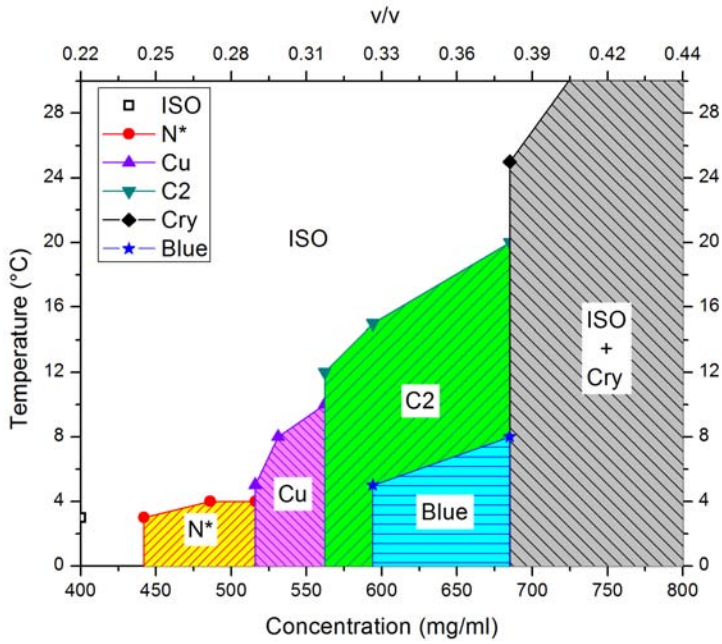


Figure 5.8: C - T phase diagram for blunt-end 5′-GTAC-3′. For increasing concentration the system orders into N^* - C_u - C_2 - Cry phases, Blue phase is found in coexistence with C_2. Increasing T causes LC melting into ISO.

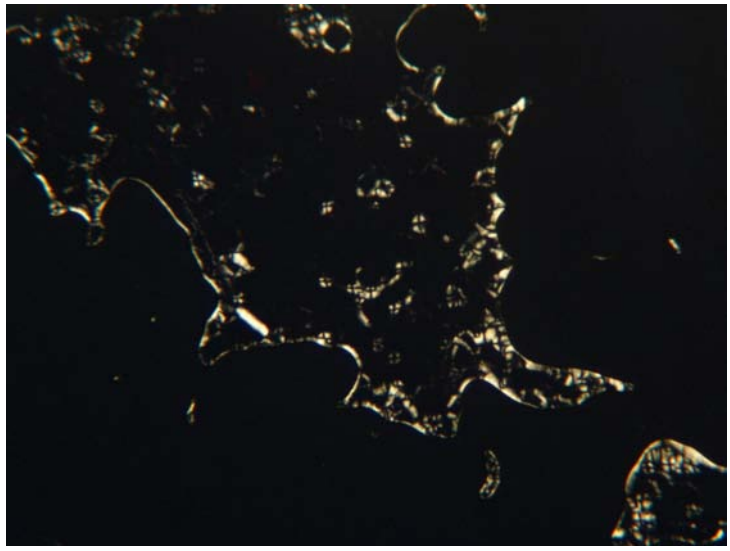
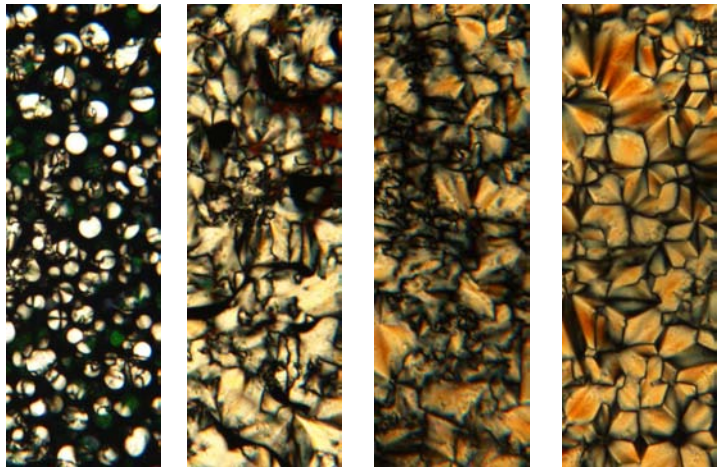


Figure 5.9: N^* phase sticky-end 5′-GCTA-3′ at $T = 1°C$. The picture has been overexposed in order to highlight the dark region that is not isotropic but long pitch cholesteric phase. It's possible to distinguish red stains with shorter cholesteric pitch. Blurred *oily streaks* are present and become thicker near the more concentrated regions at border of the LC droplet.



(a) (b) (c) (d)

Figure 5.10: LC textures of sticky-end 5′-GCTA-3′ at $c \approx 540mg/ml$. (a) Nucleation of N^* domain at $T = 15°C$ cooling from ISO, green selective reflection stains are visible; (b) N^* phase spreads all over the volume after cooling the same region at $T = 14°C$, it's noticeable the lenghtening of the cholesteric pitc; (c) transition to more crowded N^* textures at $T = 12°C$; (d) fan shaped *Focal Conics* after equilibration at $T = 4°C$

5.2.1 The granular phase of GCTA

The peculiarity of the C-T phase diagram, as reported in Figure 5.14, is that at higher temperature LCs don't melt into usual ISO phase, but into a peculiar non birefringent phase that in bright field reveals to be formed by granular shaped structures in coexistence with Isotropic. The excellent visibility of the granular structures, combined with their optical isotropy is an indication of the large mismatch in refractive index between these structures and the surrounding isotropic phase. This also makes the samples rather turbid (Figure 5.19) and suggests that DNA concentration inside the grains is bigger than outside. Doping the system with fluorescent dyes that intercalate inside DNA double helix, as EtBr, shows that grains fluorescence is higher than the surrounding part, confirming that these domains are a DNA rich phase (Figure 5.18). The size of the grains is dependent on concentration: at low concentration grains are small and dispersed, instead at increasing concentration they become denser and bigger.

Increasing temperature the granular phase slowly melts into Isotropic, T_m is in the range $40 - 80°C$. Instead, cooling from Isotropic, from $90°C$, the granular phase sharply nucleates in the temperature range $30 - 50°C$ depending on the concentration. After nucleation the domains size grows and coalescence is typically observed, but complete separation of the system into two phases has never been achieved. The system seems to fall into a dynamically arrested state, probably due to the high viscosity of the system.

Cooling even below LCs formation temperature the granular phase still remains, indicating that the two phases are different enough in structure to be competitive (Figure 5.15 and 5.17). When instead the temperature is rapidly quenched, the granular phase is

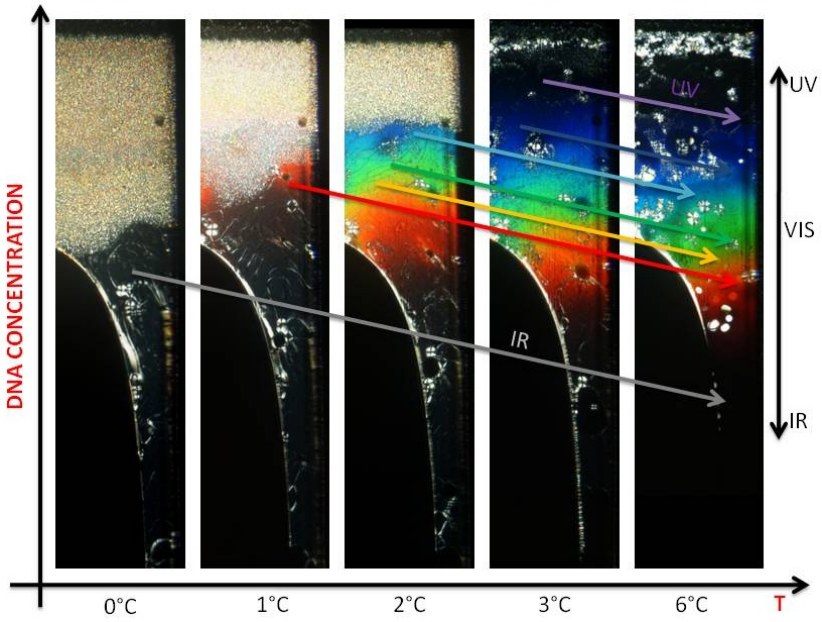


Figure 5.11: N^* phase of sticky-end 5′-GCTA-3′ evaporating in a $50\mu m$ thick capillary: rainbow effect in IR-VIS-UV region is due to variation of the selective reflection as consequence of the decreasing of the cholesteric pitch in function of concentration and temperature. Black area is the air meniscus.

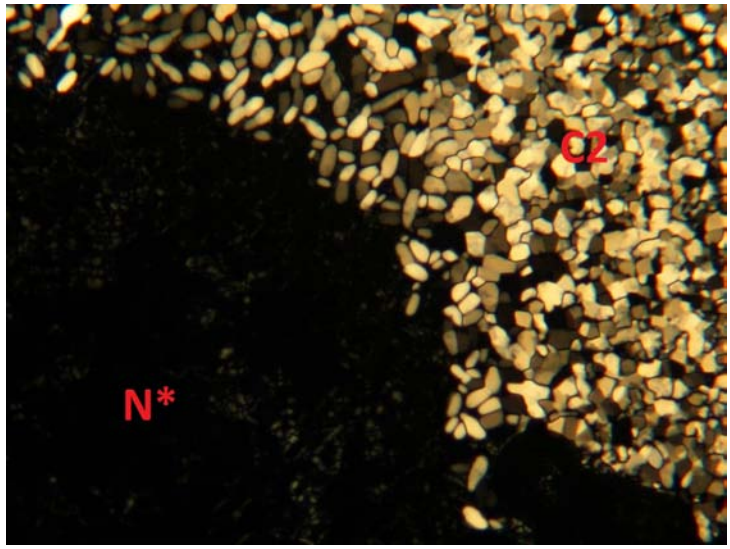
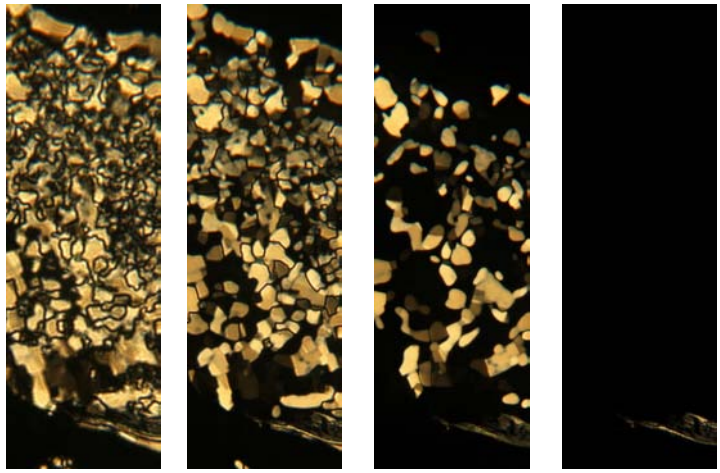


Figure 5.12: $N^* - C_2$ boundary of sticky-end 5′-GCTA-3′ at $T = 1°C$ in a concentration gradient. Please note the unusual absence of C_u phase in the intermediate region.

″skipped″ and the system only develops liquid crystal ordering (Figure 5.16).



(a) (b) (c) (d)

Figure 5.13: Melting of *Mosaic* textures of C_2 phase of sticky-end 5′-GCTA-3 at very high concentration $c \approx 750 mg/ml$: (a) $T = 12°C$; (b) $T = 26°C$; (c) ISO-C_2 at $T = 28°C$; (d) ISO at $T = 30°C$.

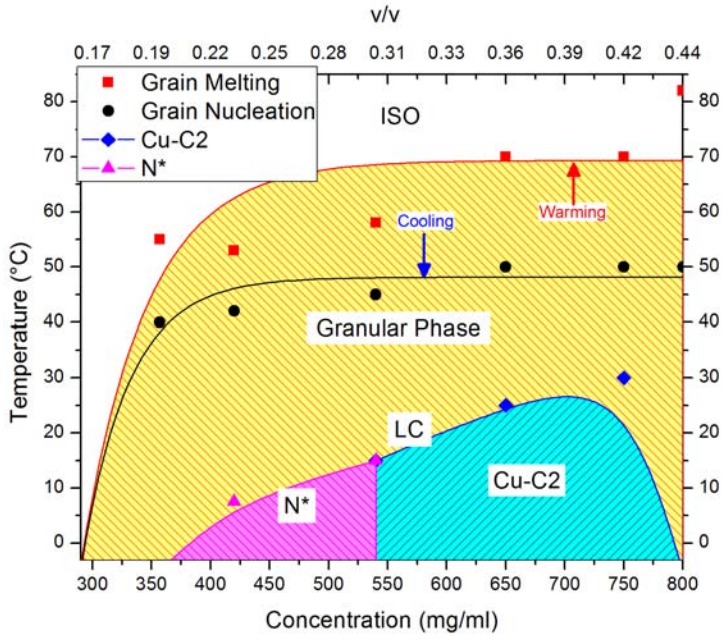
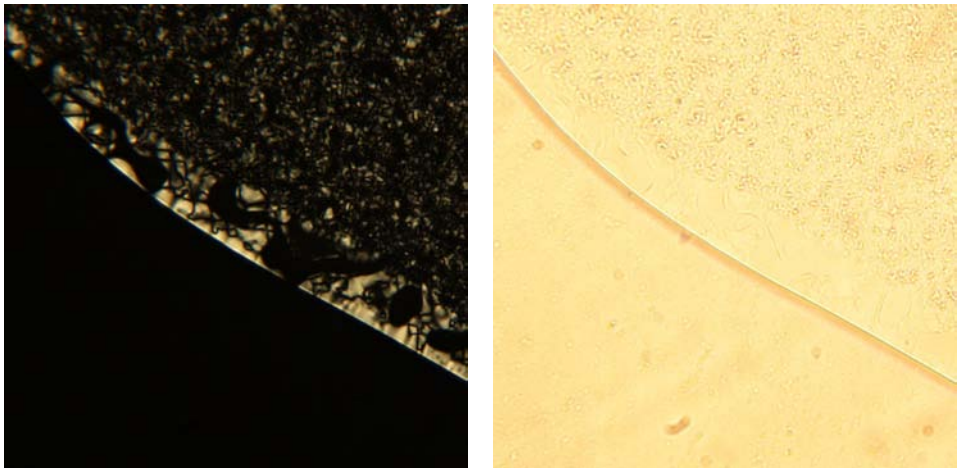
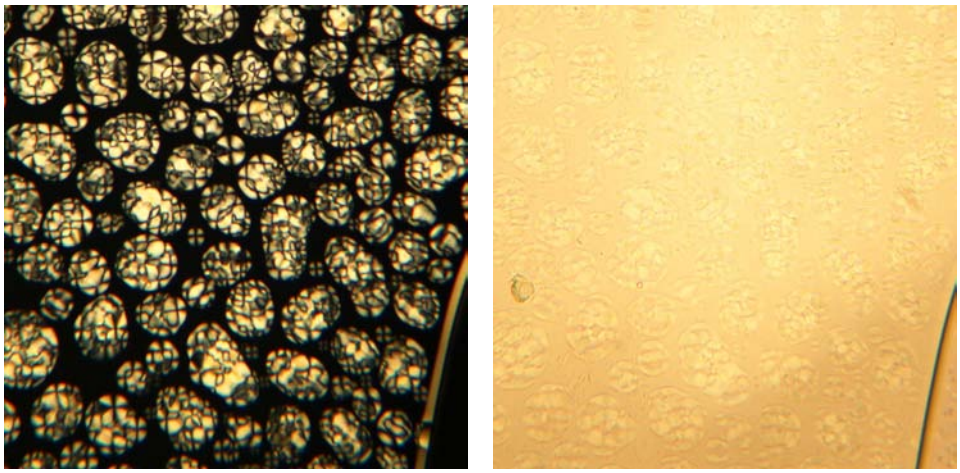


Figure 5.14: Peculiar C - T phase diagram for sticky-end 5′-GCTA-3′. At low temperature various Liquid Crystalline phases are found from N^* to C_2 as concentration increases in the range $420 - 800 mg/ml$, even if a clear C_u region is difficult to define. Increasing Temperature LCs melt into a granular non birefringent phase coexisting with Isotropic (ISO-ISO). Melting (red) and nucleation (black) temperatures are reported. Lines are the hypothetical phases boundaries.



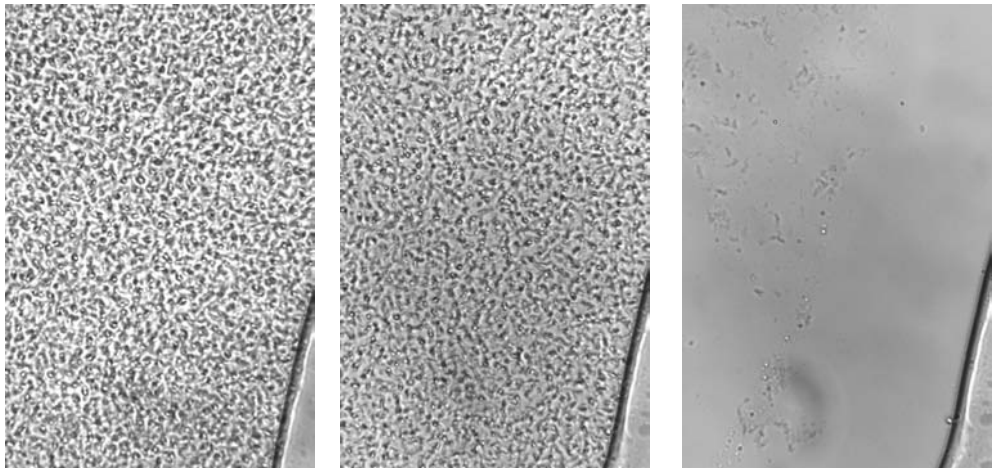
(a) (b)

Figure 5.15: Granular phase competing with N^* at $T = 5°C$: (a) polarized and (b) bright field picture of the same region.



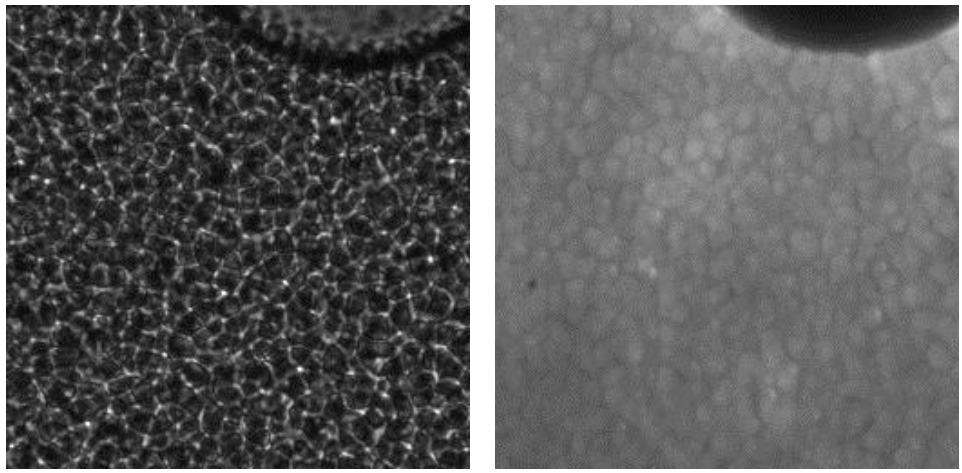
(a) (b)

Figure 5.16: Reappearance of LC phase of sticky-end 5′-GCTA-3′ after fast cooling down to $T = 5°C$ of the sample, same region of Figure 5.17: (a) polarized image of N^* domains in coexistence with ISO; (b) bright field image shows absence of granular structures.



(a) (b) (c)

Figure 5.17: Bright field images of granular phase of sticky-end 5′-GCTA-3′ after slow cooling of the sample: (a) at $T = 5^{\circ}C$, i.e. below LC temperature it prevents LC formation; (b) at $T = 40^{\circ}C$, i.e. this phase is stable up to higher temperature than LC; (c) at $T = 50^{\circ}C$, melting of the granular phase into Isotropic (unmelted structures are impurities).



(a) (b)

Figure 5.18: Granular phase of sticky-end 5′-GCTA-3′ at $T = 40^{\circ}C$: (a) bright field image, at high concentration coalescence of the granular domains is observed; (b) fluorescence image of the same region, fluorescent emission of intercalated EtBr confirms that these domains are a DNA rich phase.

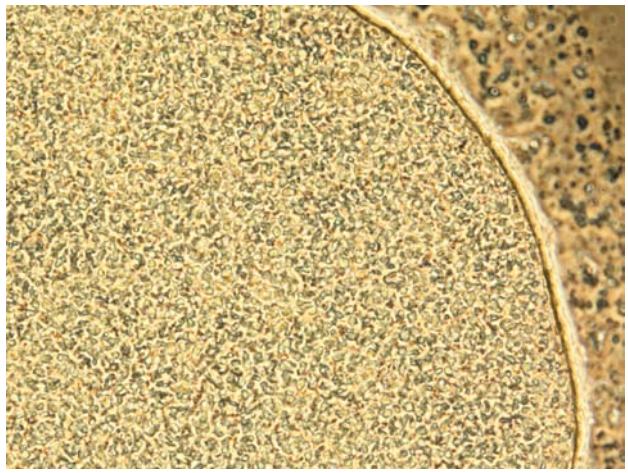


Figure 5.19: Phase contrast image of granular phase of sticky-end 5′-GCTA-3′, considerable difference in refractive index between grains and background is evident.

5.2.2 Concentration and nature of the granular phase

In order to estimate the difference in concentration between the granular phase and the Isotropic background, samples at different DNA concentration has been studied. Sample A at $C_A = 360mg/ml$, sample B at $C_B = 410mg/ml$, sample C at $C_C = 540mg/ml$, sample D at $C_D = 650mg/ml$ and sample E at $C_E = 650mg/ml$. The fraction of the granular phase has been extracted for these samples by analysis of bright field images (Figure 5.20). The fraction of the granular phase increases with the concentration of the samples: $\Phi_A = 5.5\pm0.5\%$, $\Phi_B = 24\pm4\%$, $\Phi_C = 45\pm2\%$, $\Phi_D = 66\pm4\%$ and $\Phi_E = 91\pm3\%$.

Considering that the total DNA concentration is the average of the Isotropic concentration C_I and of the granular phase C_G weighted on the relative fractions, it obviously follow the equation:

$$C_{tot} = C_G \cdot \Phi_G + C_I \cdot (1 - \Phi_G) \tag{5.1}$$

from which follow:

$$C_{tot} = \Phi_G \cdot (C_G - C_I) + C_I \tag{5.2}$$

Plotting C_{tot} vs Φ_G and linear fitting the data (Figure 5.21), it's possible to extrapolate the concentration of the two phases: the intercept gives $C_I = 310\pm20\ mg/ml$ and the sum between the slope and the intercept gives $C_G = 820 \pm 50\ mg/ml$. This estimate has been produced with the assumption that the dense phase is compact. Should we find that the collapsed DNA forms microscopic sponge-type structures due to kinetic arrest, we wold have overestimated their volume, i.e. underestimated their concentration. Sticky-end 5′-GCTA-3′ solutions at concentration lower than C_I show only the Isotropic phase confirming the lower boundary.

Although we haven't yet determined the microscopic structure of the granular phase,

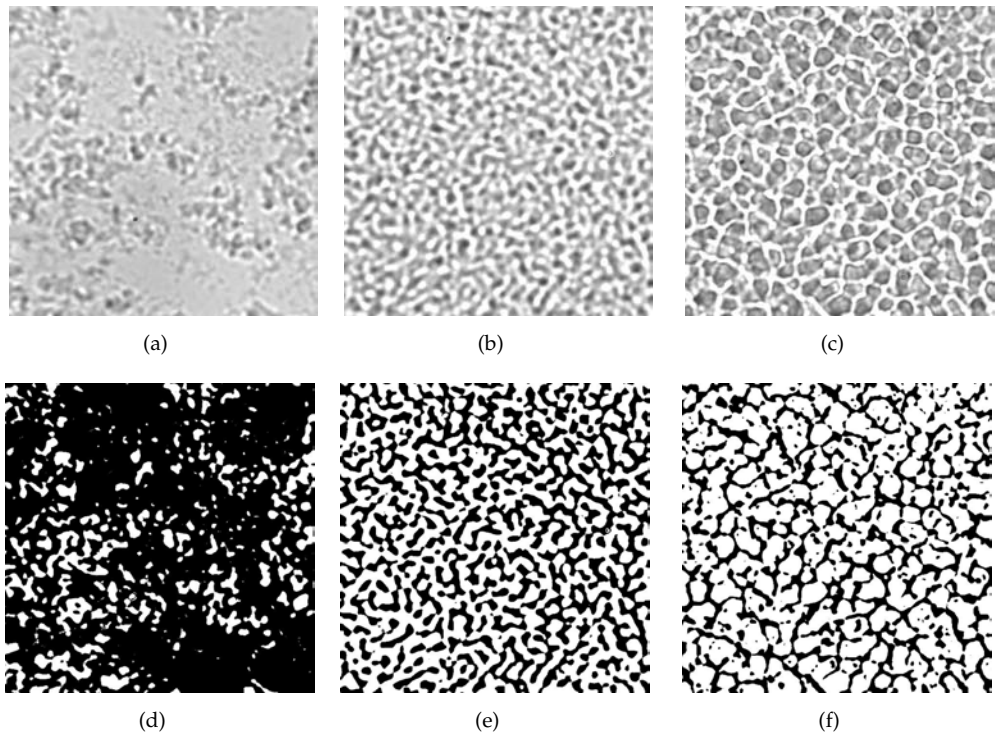


Figure 5.20: Bright field images of granular phase from samples at different concentration, analysis of the same images with intensity threshold allows to estimate the fraction of the granular phase. (a) sample B $C_B = 410mg/ml$, (b) sample C $C_C = 540mg/ml$, (c) sample D $C_D = 650mg/ml$, (d) $\Phi_B = 24 \pm 4\%$, (e) $\Phi_C = 45 \pm 2\%$, (f) $\Phi_D = 66 \pm 4\%$

the basic observations here reported indicate that it is not a consequence of linear aggregation, but rather of some kind of oligo arrangement that leads to branched structures:

- the observed globuli are optically isotropic, while linear aggregates favor parallel packing;

- the granular phase and the LC phase are metastable at the same concentration and T, indicating that they cannot evolve continuously form one another. Moreover, if the cooling process was slow enough to allow for the formation of a few globular domains, they remain even after the surrounding volume is filled by LC, indicating that even in the presence of nuclei of the other phase, the two phase don't evolve one into the other, in turn strengthening the notion that their microscopic arrangement is significantly different;

- linearly aggregating monomers, which can be modelled as di-valent monomers, are not expected to produce phase liquid-vapor-like phase separation [89]. Indeed linear DNA structures do not have the mutual attraction forces to drive their collapse into globules.

Therefore, the microscopic structure of the globules ought to involve branched aggregation, of yet undertermined type.

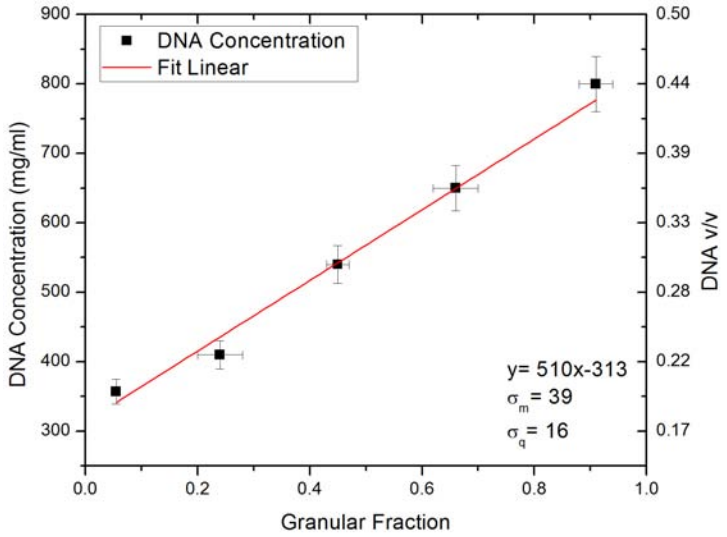


Figure 5.21: C_{tot} vs Φ_G plot for sticky-end 5'-GCTA-3' samples at $T = 40°C$. Error bars for Φ_G are statistical errors from images analysis, for C_{tot} are statistical errors from concentration measurements. The slope and the intercept of the linear fit give the concentration of the granular and Isotropic phases: $C_G = 820 \pm 50\ mg/ml$ and $C_I = 310 \pm 20\ mg/ml$.

5.3 Thermal stability of 4 bases DNA strands

Experimental thermal denaturation curves have been obtained for samples showing LC phase at room temperature, the collected data revealed an incredibly high melting temperature for either blunt-ends $T_M \approx 50°C$ and, even higher, for sticky-ends $T_M \approx 60°C$ (Figures 5.22 and 5.23).

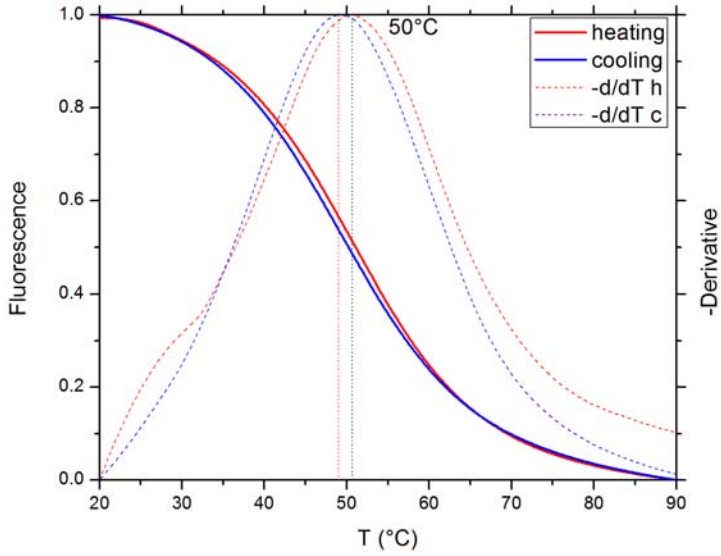


Figure 5.22: Experimental thermal denaturation of blunt-ends 5′-GTAC-3′. Normalized EtBr fluorescence curves (solid) and their negative derivative (dash): T_M is the point of maximum negative slope, i.e. the maximum of the negative derivative.

The free energy gained in the pairing of two 4mers is rather small, resulting in a low melting temperature, much lower than the one observed in our measurements. For sure the NN approach and the values in the database are not meant to be accurate for such short oligonucleotides and for such high concentrations. However, it is interesting to note that there are elements that might help understanding why the melting T is actually so much larger than expected. One effect certainly comes from the number density of oligomers. Given the low molecular weight of the 4mers, the concentrations at which liquid crystals phases are found correspond to a high molar concentrations, which become extremely large for the shortest oligomers. Indeed, a concentration of $600mg/ml$ corresponds to a molar concentration of $0.16M$ of 12mers and to $0.5M$ of 4mer oligos. It is interesting, from this standpoint, to compare the melting T expected for blunt-end oligomers at (normal) low concentration to the melting T at $600mg/ml$, as a function of oligomer length (Figure 5.24). Quite clearly the effect of the high molar concentration is very relevant, especially for the shortest oligos.

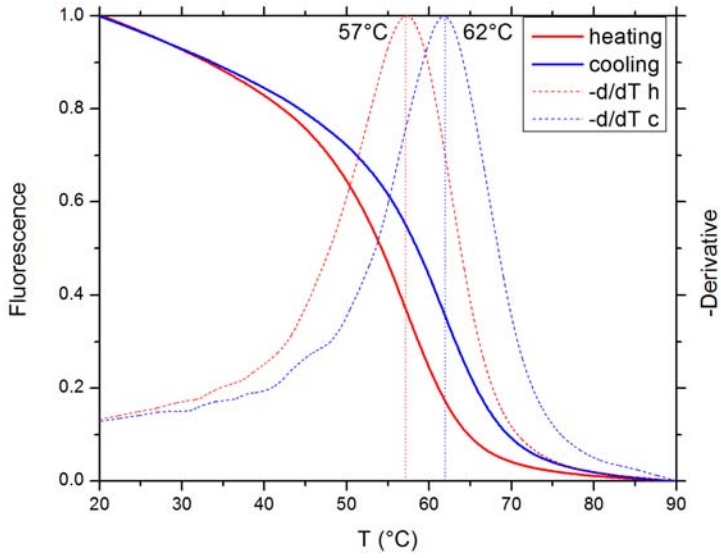


Figure 5.23: Experimental thermal denaturation of sticky-ends 5′-GCTA-3′. Normalized EtBr fluorescence curves (solid) and their negative derivative (dash): T_M is the point of maximum negative slope, i.e. the maximum of the negative derivative.

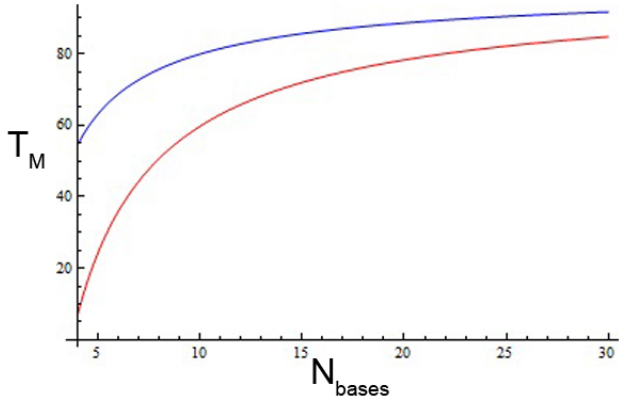


Figure 5.24: Plot of the calculated T_m vs DNA strand length at two different oligomers concentration: $1mg/ml$ (red) and $600mg/ml$ (blue). The effect of the high molar concentration is very relevant, especially for the shortest oligos.

5.3.1 Sticky-end linear aggregation

The rationale for this unusual behaviour can be found in the fact that for oligomer constituted only of 4 bases 5′-GCTA-3′, that hybridizes forming a duplex with 2 paired core bases and sticky tails of 2 bases each, the Gibbs free energy gain of the core pairing is close to that of the sticky tails. In such a scenario a cooperative pairing between more than two strands is more favorable than the formation of a single duplex. For example: the formation of the duplex with GC core has $\Delta G_{GC} = -3.14kcal/mol$, the addition of

another strands would add two base pairs more and a coaxial stacking stabilization that increases the Gibbs free energy to $\Delta G_{3strands} = -5.7kcal/mol$, the addition of a forth strand would give $\Delta G_{4strands} = -10.44kcal/mol$, and so on. Thus, after the nucleation of a duplex, the system has more convenience in evolving by growth of a longer chain by addiction of single strands than in seeding other new duplexes.

This is the main difference between this system and those constituted by longer oligomers, as 14mer with a 12 bases core and 2 bases sticky tails, where the ΔG of the duplex formation ($\approx -29kcal/mol$) is much more than that of the tails interaction $-2.9kcal/mol \leq \Delta G_{sticky} \leq -3.6kcal/mol$ (data from [44]), leading to two separate steps: first the formation of duplexes, at higher temperature, and only after that the growth of linear chains by tails interaction, at lower temperature.

5.3.2 Blunt-end linear aggregation

Completely self-complementary 5′-GTAC-3′ strands instead hybridize forming blunt-ends duplexes which interact only through aromatic stacking of the terminal bases. GC terminal bases have a relatively high interaction energy, $\Delta G = -3.1kcal/mol$ (From [44]) that is comparable with the hybridization energy, $\Delta G = -3.65kcal/mol$. At high concentration this interaction promotes the linear aggregation of nicked duplexes into longer chains, improving in this way also the duplexes stability.

The self-assembly of 4-base-long oligomers takes place in a different regime with respect to the liquid crystal ordering of longer oligos. The very fact that the energies involved in the formation of the pair are of the same order of the energies involved in the linear aggregation implies a change of perspective: aggregation and duplex denaturation should be considered as a single event. As a consequence, the calculation of the melting temperature of the duplex should include the contributions coming from the other aggregated duplexes. In other words, one should not ask the temperature at which a duplex melt, but rather the temperature at which a single oligomer chain detaches from an aggregated chain. Accordingly, in the calculation of T_m the energetic contribution from the stacking or coaxial stacking with adjacent chains has to be taken into account. Figure 5.25 shows the calculated T_M if it is defined as the temperature at which a single oligomer chain detaches from an aggregated chain.

5.4 Additional observations

Here we report other observations that confirm the variety and robustness of the phenomena described and provide ideas for future research directions.

- Sticky-end 4mer with sequence 5′-ATTA-3′ shows LC ordering. Cholesteric and Columnar phases have been found at concentration $\approx 630mg/ml$ and temperature $-3°C < T < 16°C$, i.e. similar conditions to those of the other 4mers more largely investigated in this work (figure 5.26a,b).

- Mixtures of sticky-end 5′-GCTA-3′ and blunt-end 5′-GTAC-3′ show coexistence of the granular phase, typical of sticky duplexes, with low birefringence mosaic tex-

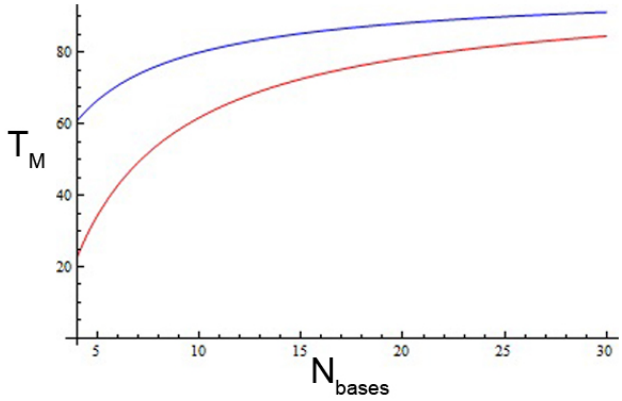


Figure 5.25: Plot the calculated of T_m vs DNA strand length, taking into account the energetic contribution from the stacking or coaxial stacking with adjacent chains. Two DNA concentration are reported: $1mg/ml$ (red) and $600mg/ml$ (blue).

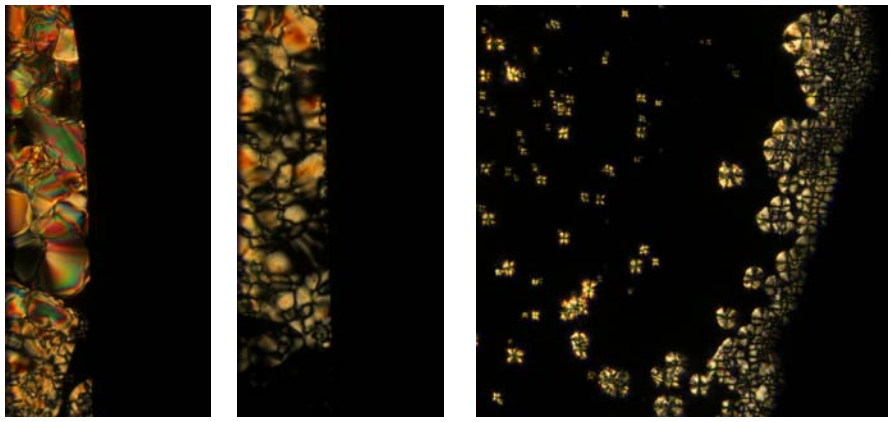
tures, typical of blunt duplexes: thus suggesting the peculiar tendency of these molecular shapes to demix at high concentration (figure 5.27).

- The preliminary finding of birefringent textures in solutions of sticky-ends 3mers, 5′-GCA-3′ and 5′-GCT-3′, that pairs with 2 core bases and 1 base sticky tails, at very low temperature is a promising approach in order to push downward the length lower limit of LC forming oligomers (figure 5.26c).

5.5 Conclusions and outlook

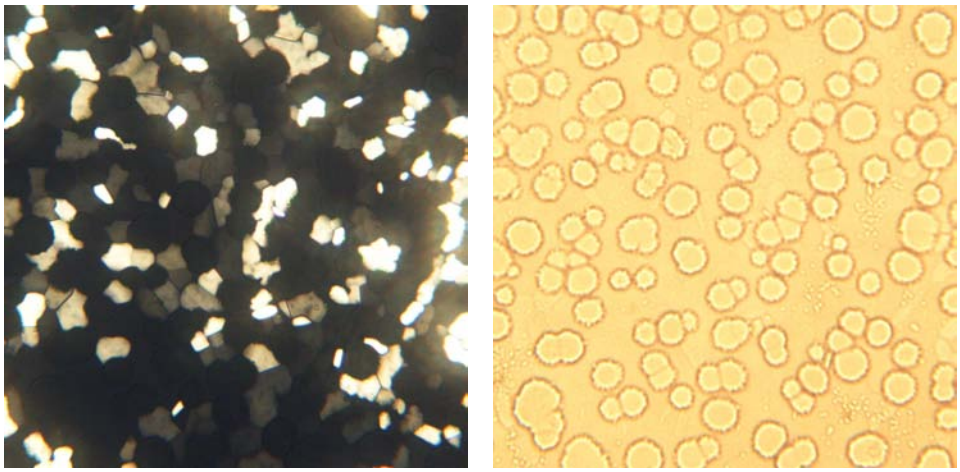
Liquid Crystalline ordering of 4 bases DNA oligonucleotides has been observed and reported for the first time. Identification of Cholesteric, Columnar and Crystal phases shows the continuity of these phenomena with already known behaviour of short DNA oligomers at high concentration. On the other hand several observations of condensed phases of such ultra-short DNA strands show peculiarities, such as Blue phase like textures and granular ISO-ISO phase coexistence, which will require more investigations in future. The collected data on thermal stability of both duplexes and LC phases suggest that the aggregation mechanism is slightly different from what has been already reported in literature for short DNA duplexes in the range $6 - 20$ base pairs.

The distinctive properties of objects possessing intra-molecular pairing interaction in the same energy range of the inter-molecular aggregation interaction open a very promising new investigation branch in the field of nucleic acids research, with possible results of general interest for soft and living matter. Moreover, pushing downward the minimum oligomers length for spontaneous formation of ordered phases, these observations add relevance to the self-assembly of short DNA strands as possible route for nucleic acids formation in primordial Earth.



(a) (b) (c)

Figure 5.26: (a) Columnar Uniaxial phase of $5'-ATTA-3'$; (b) Cholesteric phase of $5'-ATTA-3'$; (c) Birefringent textures of 3mers 5′-GCA-3′ and 5′-GCT-3′.



(a) (b)

Figure 5.27: Coexistence of granular phase and bluish phase with low birefringent mosaic textures in blunt-sticky mixtures. (a) polarized image and (b) bright field image of the same region are showing the tendency of the system to demix.

CHAPTER 6

Ligation of short DNA oligomers in LC phase

The emergence of liquid crystal ordering of DNA and RNA oligomers, as reported in the previous chapters, enlightens a remarkable and finely tuned combination of staged self-assembly: DNA hybridization, linear aggregation, liquid crystallization, phase separation and compartmentation of sequences, non-equilibrium effects. Indeed the stacking of duplexes within the LC phases, in which the terminals of the oligonucleotides are held in continuous physical contact, appears as a templating environment that could favor the selective non-enzymatic ligation of short complementary oligomers into longer complementary ones, a crucial step in the appearance of early life [90].

The aim of this part is to test the hypothesis that self-assembly into LC phases influences polymerization of DNA double strands under conditions in which chemical ligation is induced. For this purpose my research is not focused in finding prebiotically plausible chemical reactions, but rather in inducing an adequate reaction in LC phases and in probing whether or not LC self-assembly gives positive contributions in reaction yields and in the quality of the emerging products.

6.1 Self-assembly as a guiding hand for linear polymerization

In confronting the prebiotic world, one of the greatest challenge is the appearance of nucleic acid polymers. As the Nobel laureate C. De Duve commented, "How RNA could possibly have emerged from the clutter without a "guiding hand" would baffle any chemist. [...] The need seems inescapable for some autocatalytic process such that each lengthening step favors subsequent lengthening. Only in this way could the enormous kinetic obstacle to chain elongation be surmounted. [...] Any invoked catalytic mechanism must accommodate the participation of a template, for there can have been no emergence of true RNA molecules without replication". [6]

We propose here a scenario in which a positive feedback between self-assembly and non-enzymatic ligation guides the elongation of short fragments of DNA and RNA to form chains having the lengths required for expressing enzymatic activity.

The conjecture here proposed is based on the features emerged in the study of LC ordering of short DNA:

1. **Self-templating**

The hierarchical self-assembly of complementary or partially complementary short DNA oliogonucleotides results in the appearance of LC phases. This phenomenon, previously reported for duplexes of minimum length of 6 base pairs, is here extended down to 4 base pairs, as discussed below. LC ordering is promoted by the linear aggregation resulting from the stacking of the duplex-forming oligomers. Conversely, the formation of LC ordering greatly enhances the stability of such chemically discontinuous but physically continuous aggregates [35, 46, 47]. Within the LC structure, duplex terminals are held in physical contact. It was recently shown that such proximity becomes a constraint within $5\mathring{A}$ when the oligomers terminate, at one of their extremities, with a phosphate [37].

2. **Self-sorting**

   LC formation is selective on the molecules that participate to the ordered structure. To a certain extent, this is a general feature in the phase diagram of mixtures. The onset of ordered phases strengthens the constraints on the compatibility of molecules, and often results in specific partitions of the molecules among the phases. This phenomenon is particularly striking in the particular case of DNA LC, where oligomers that form duplexes able to interact end-to-end tend to separate from single stranded oligomers or oligomers with defects impairing them to aggregate. This phenomenon effectively acts as a molecular sorter favoring the spatial localization of aggregating oligomers into microsized LC domains. The mechanisms behind this behavior are both enthalpic and entropic, and are found to be active also when duplex-forming and mutually interacting oligomers are mixed with other molecules other than nucleic acids, such as PEG.

3. **Non-enzymatic ligation**

   The physical proximity of the terminal groups and the symmetry of LC packing favoring linear structures become decisive when conditions are such to favor the covalent chemical ligation of the oligomer terminals. We have previously speculated that when such conditions are achieved, LC ordering may have a significant effect on the efficiency and on the structure of the resulting chains. Here we prove such conjecture by studying the oligomers resulting form non-enzymatic ligation in a few paradigmatic conditions. Oligomer elongation through ligation stabilize the LC ordering, extending the temperature and concentration ranges where it can be found, and provide nuclei in turn granting a faster growth of LC ordering. Thus, LC-templated oligomer elongation and self assembly-based selection can operate as part of a positive feedback loop in the presence of ligation wherein modes of ligation which stabilize the self–assembly are promoted or favored.

6.2 DNA oligomers condensation with carbodiimides

6.2.1 EDC as zero length cross-linker

One of the most efficient methods for chemical linkage of oligonucleotides, via $5' - 3'$ phosphodiester bond, is condensation with carbodiimides as water-soluble 1-ethyl-3-(3-

dimethylaminopropyl) carbodiimide (EDC). Carbodiimides are probably the most popular type of zero length cross-linker in use nowadays [91]. Zero length means that the cross-linker does not add any extra atoms to the two molecules to be conjugated, and those form a simple covalent bond without any added spacer, i.e. the product of the linkage of two oligonucleotides is a longer strand that effectively is a totally natural DNA molecule.

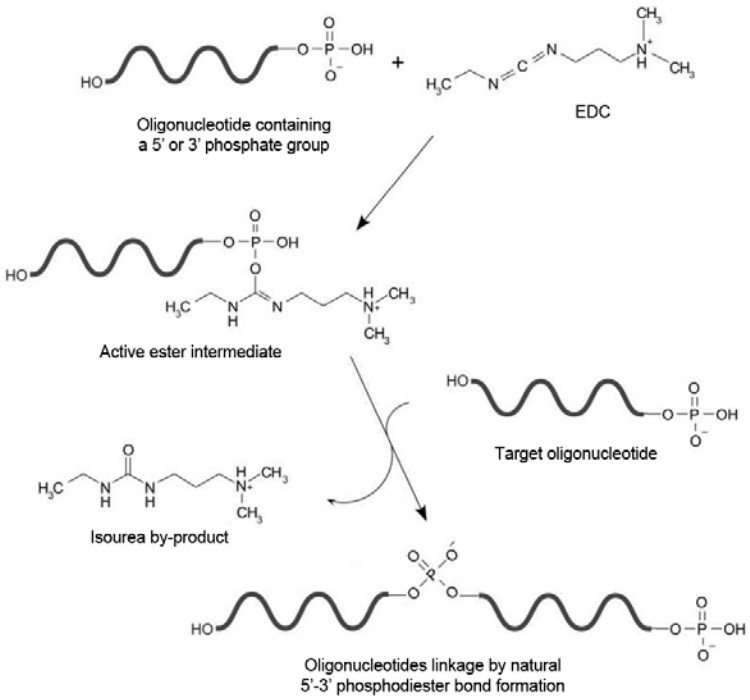


Figure 6.1: Scheme of EDC activated ligation of two oligonucleotides: oligonucleotides containing a terminal phosphate group can be reacted with EDC to form an active intermediate. This derivative is highly reactive with nucleophiles as OH termination of a second DNA strand, forming a natural phosphodiester linkage. (modified from [91])

In presence of nucleic acids (fig. 6.1), EDC efficiently reacts with a terminal alkyl phosphate forming an intermediate complex, that can highly react with a nucleophile, like a primary amine ($-NH_2$), to form a phosphoramidate bond, or with an hydroxyl ($-OH$), to form a phosphodiester bond. In aqueous solution the reaction has to compete with hydrolysis by water, forming isourea as side product, and restoring the original phosphate group. Therefore this competition mostly affects hydroxyls, whose nucleophilicity is closer to the one of water oxygen atoms: the target molecule would have comparable reactivity to that of any free OH of the aqueous environment. Instead, due to their higher nucleophilicity, amine-containing compounds are preferred target for this kind of reaction, allowing higher rate, yield and specificity. Each reaction step is highly dependent on pH. Acidic pH values favor EDC hydrolysis, with increasing stability of

the carbodiimide in solution at or above pH 6.5. The protonation of EDC is also important in the formation of the reactive intermediate. Phosphate di-anions are good nucleophiles, while phosphate mono-anions are poor nucleophiles. Most references to the use of EDC describe the optimal reaction medium to be at a pH from 4.7 to 6.0. However, the carbodiimide reaction occurs effectively up to at least pH 7.5 without significant loss of yield.

6.2.2 Use of EDC in ligation reactions

During the 80s Shabarova and co-workers started developing non enzymatic approaches to obtain completely chemical linkage of DNA duplexes. They called this method chemical "ligation" and it consists of the condensation of DNA strands in presence of the complementary sequence as template and with the use of chemical activators of the terminal phosphate, like water-soluble carbodiimides (such as EDC) [92] or cyanogen bromide (CNBr) [93]. Those condensing reagents have different composition and reaction kinetics. In [92], Shabarova and coworkers obtained polymerization of the concatemeric duplex formed by asymmetrically overlapping parts of the 3′- or 5′-phosphorylated decanucleotide 5′-TGGCCAAGCT-3′ in presence of EDC. As product $20 - 100$ bases long DNA strands were containing only natural phospodiester bonds with 90% reaction yield in optimal conditions. Activated 3′-phosphates (3′P) were shown to be significantly more reactive than 5′-phospate ones (5′P). Moreover both products yield and reaction rates depend on the local structure of the double helix at the bonding site: the chemical properties of 3′ and 5′ termination of the nicked building blocks have a great impact on the efficiency of the ligation reaction, like amino (NH_2) substitution of the natural hydroxy-group OH. According to [94], the natural phosphodiester bonding reaches yield of 95% in 4 days for $5'OH - 3'P$ nicked strands and 75% in 6 days for $5'P - 3'OH$; instead $5'P - 3'NH_2$ nicked strands have 97% yield in 6 hours with EDC. The use of CNBr gives respectively yields of 95%, 35%, 86% in 1 minute reaction. Even if CNBr is highly reactive, its ligation products have generally lower yield in comparison to those of EDC. Table 6.1 compares the relative reactivity of EDC reaction with combination of different terminals of the nicked DNA strands. The great performances of reactions involving amino terminated oligonucleotides is due to the higher nucleophilicity of NH_2 group that reacts up to 100 times faster than natural hydroxyl with EDC activated phosphate. For this reason this approach has been widely used for non enzymatic ligation reactions. For example self-complementary deoxy-dinucleoside $5'P - GC - NH_2 - 3'$ is found to oligomerize via EDC condensed phosphoramidate bonds up to 15 times the initial length, with a yield of more than 70% [95].

6.2.3 Use of EDC in non enzymatic autocatalytic reproduction

Many efforts have been made in searching autocatalytic reproduction of DNA/RNA strands in absence of enzymes, i.e. to find a chemical system that favor oligonucleotide self-replication. For this purpose the original strand has to be composed of both purine and pyrimidines arranged in a self complementary sequence, in such a way to catalyze the reproduction of an identical strand. Moreover the ligated product has to sep-

3′	5′	Relative Reactivity*	Relative Reactivity**
OH	P	1	1
P	OH	2	10
P	P	29	45
NH_2	P	31	70
P	NH_2	-	100

Table 6.1: Relative Reactivity of EDC reaction with combination of different terminals of the nicked DNA strands. * data extracted from [94], ** from [96]

arate from the template, in order for the system to evolve, i.e. temperature changes are needed. Von Kiedrowski [97] obtained the first nonenzymatic self-replicating system by the autocatalytic ligation of two trinucleotides (A and B) on a hexameric self-complementary template (T) (fig. 6.2). Ad hoc terminal protection allows water-soluble 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide to activate only 3′-phosphate of strand A, and then either the formation of natural phosphodiester bond with trimer B or pyrophosphate bondig with another A strand. Reactions have been performed with increasing template concentration at both $0°C$ and $20°C$. While at the latter pyrophosphate formation takes place and template effects are not relevant, at low temperature natural hexamer replication increase with increasing template concentration. Thus autocatalytic replication is found only if temperature is low enough to allow hybridization of trimers A and B on hexamer T, i.e. a template driven condensation. Reaction conditions are: 12 mM trimer A, 10 mM trimer B, 0 - 0.2 - 0.4 - 0.8 mM template concentration.

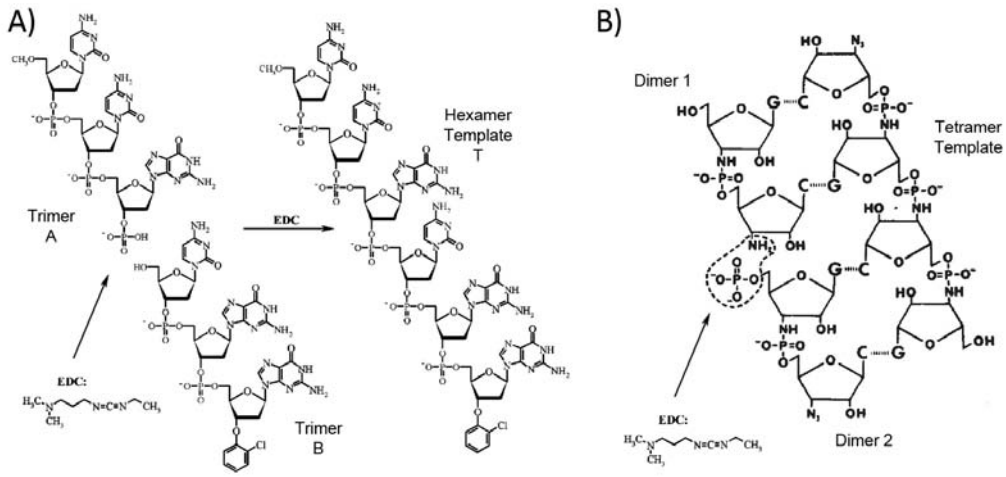


Figure 6.2: Autocatalityc synthesis of an hexanucleotide (A) from [97] and of a tetranucleoside phosporamidate (B) from [98]. Two nicked strands are held in proximity by pairing with a template sequence, EDC condesation leads the formation of a new strand that is identical to the starting template.

Zielinski and Orgel [98] achieved the autocatalytic synthesis of a tetranucleotide ana-

logue from dimers complementary to a tetranucleoside triphosphoramidate template in presence of water-soluble carbodiimide. The system (fig 6.2) is designed in order to allow only the phosphoramidate linkage between 5′-phosphate and 3′-amine of two nicked dimers on a complementary tetramer. Other terminal groups are ″capped″ in order to eliminate side reactions different from the template copy. The yield of the reproduced strands increases with the square root of the template concentration. Reaction conditions are: 1 mM for each dimer, 60 mM EDC, 0.25 - 0.5 - 1 mM tetranucleoside triphosphoramidate template.

6.2.4 Recent use of EDC in DNA ligation reactions

Recently, in 2010, Von Kiedrowski group realized the non enzimatic polymerization of long DNA strands from short double stranded deoxynucleotides [7]. They showed that 5′-amino-3′-phospate pyrimidine-purine dinucleotides, tetranucleotides and hexanucleotides activated by EDC oligomerize giving rise to longer strands with 3′-5′ phosporamidate linkage. The complementarity of the reacting building blocks is the key role that prevents cyclization and allows the efficient polymerization of long products, up to 300 bases.

After 72 hours of reaction time the estimated DNA strands final length doesn't seem to depend neither on the starting oligonucleotide's characteristics as length, sequence, self-complementarity, nor, when different building blocks are mixed, on the mixture composition, like presence of template strands or double helix pairing energy. Instead the analysis of the reaction rates shows a high dependence on the inter-duplex stacking interaction: CG terminated building blocks react faster than AT terminated ones at fixed starting length. For equally terminated duplexes the rates increase with the starting length of the building blocks.

Polycondensation reactions have been performed in 0.1 M HEPES buffer pH 7.5 at $2°C$, mostly at 20 mM oligonucleotides concentration. Changes from 5 to 50 mM in starting concentration doesn't significantly influence neither the maximum length of the products after 48 hours nor the reaction rates as it would be expected in a second order polycondensation reaction, i.e. dependence on the square power of the reactants concentration. This means that the reaction rate is not limited by bimolecular polycondensation but by supramolecular aggregation of the reactants that precedes ligation.

In her PhD thesis, Taran [96] distinguishes two polymerization regimes. The first is a non-cooperative (*isodesmic*) aggregation, in which association of the monomers is independent on the forming aggregate length. Tetranucleotides and hexanucleotides building blocks are found to follow this aggregation path. Indeed for longer self complementary oligomers faster reaction rates are due to the fact that at fixed molar concentration the volume fraction increase with helices length, which favors stacking and this way supramolecular aggregation.

The second is a *cooperative* aggregation, in which the association constant of the monomer is dependent on the forming aggregate length. Within this model the formation of the first aggregation nuclei is slow. This is the case of dinucleotides, whose stacking interaction is unfavored because of the monomers dilution. Their initial aggregation is instead

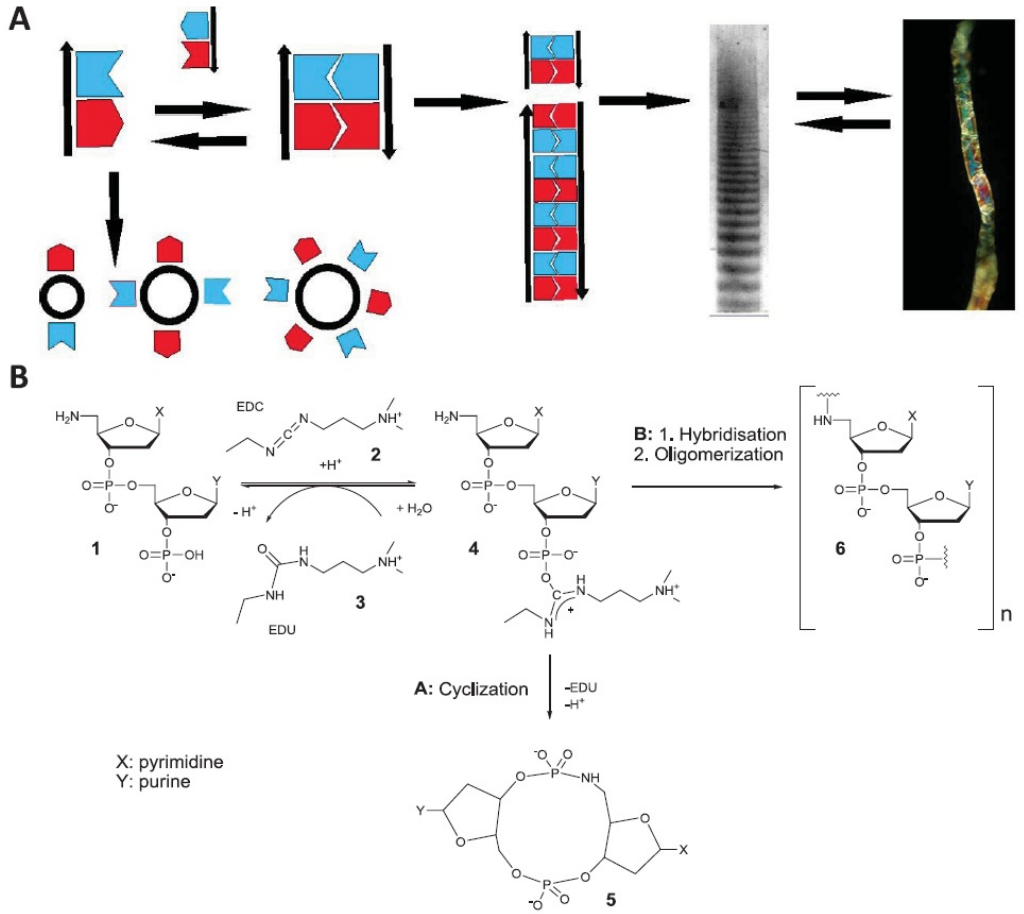


Figure 6.3: EDC activated polycondensation from [7] (A) Complementarity of the building blocks promote chain elongation vs cyclization. (B) EDC activate 3′ Phosphate forming a highly reactive group with good nucleophiles as 5′-amino termination or water OH.

favored by electrostatic attraction: at pH 7.5 3′ terminal phosphate is partially deprotonated having charge -1 or -2, EDC activation changes terminal phosphate charge to $+1$, thus changing the total charge of the strand, that for a dinucleotides means to switch from negative to positive. This causes an increasing in attraction between dinucleotides building blocks at the beginning of the reaction, forming longer strands that will start to aggregate, and thus leading a cooperative behavior.

Changes in EDC concentration from 0.05 M to 1 M cause variation in the reaction rate but the maximum products length doesn't show appreciable changes. The tendency toward the saturation of the product length, in my opinion, can be due to the large amount of EDC used in these experiments, that overcome the concentration of the building blocks. Secondly it can be a consequence of the statistical distribution of the aggregates lengths. Moreover significant differences can be hidden by the rather weak resolution of polyacrilamide gels at high lengths. Nevertheless the product distribution

seems to shift slightly towards longer strands at increasing EDC concentration.

6.3 Planning of the experimental system

The aim of this study is to test the hypothesis that LC order influences polymerization of DNA double strands under conditions in which chemical ligation is induced. For this purpose we focused on self-complementary oligonucleotides already well characterized in recent works [1],[44] and reported in table 6.2, with addiction of phosphate at one terminal and keeping the natural hydroxyl on the other. Both 5′ and 3′ phosphate have been tested.

Sticky-ends
First experiments have been done with sticky-end 14mer 5′-CGCGAATTCGCGCGp-3′ (DD-CGp), MW 4261 Dalton, that self-hybridizes forming a 12 base pairs core helix with 2 bases long sticky tails at both ends. Pairing interaction between CG tails allows duplexes to polymerize into linear aggregates, maintaining natural DNA 5′-3′ direction.

..... 5′-CGCGAATTCGCG-**CGp**/CGCGAATTCGCG-**CGp**-3′ ...
... 3′-**pGC**-GCGCTTAAGCGC/**pGC**-GCGCTTAAGCGC-5′ .....

The peculiar geometry of this kind of system constrains 3′ terminal phosphate to stay in proximity to the adjacent 5′ hydroxyl, thus in the optimal position for inter-strand linkage. Moreover DD-CGp shows liquid crystal ordering at relatively low concentration (400-500 mg/ml). For these reasons it's a good testing system for EDC condensation.

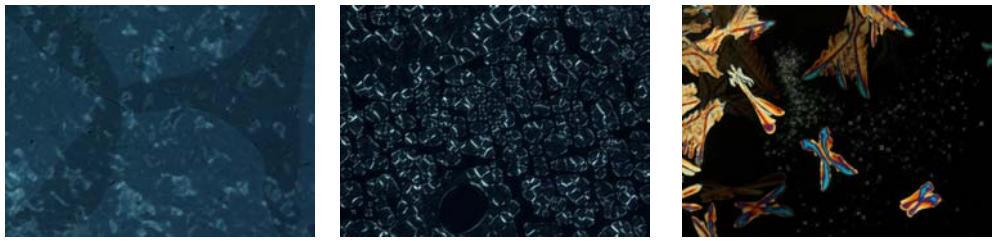
Blunt-ends
A more challenging task is to induce ligation in a blunt-end system because helices interact only by end-to-end stacking force, mostly mediated by the aromatic stacking of the terminal bases, that has lower intensity than the sticky-end interaction. For this reason, see [35, 44, 46, 47], the aggregation degree in this case is expected to be lower than with sticky-end building blocks, corresponding also to a higher concentration range for LC formation.

Moreover blunt end-to-end interaction has less geometrical constraints than sticky-end, i.e. in principle two stacked helices are free to rotate around their vertical axis. This is not guarantying the natural helical continuity to be steadily maintained. This topic has been object of recent studies and computer simulations [37]. It is found that terminal phosphates, negatively charged, tend to minimize repulsion by leading the helix twist to a preferential angle of $-20°$. This conformation has $\Delta\Delta G = -2.2kcal/mol$ respect the natural intra-strand DNA twisting angle that is $36°$. It has to be taken into account also the fact that, in this system, EDC-phosphate binding changes the total charge of the terminal group from negative to positive, thus probably modifying final twist angle.

Best candidate for these kind of experiments is blunt-end 12mer 5′-CGCGAATTCGCG-3′, known as "Dickerson Dodecamer" (DD) already well studied for its interesting properties, like self-complementarity, equal number of purine and pyrimidine bases, stability of the CG terminals.

Sequence	$C_{IN}(mg/ml)$	$C_{NC}(mg/ml)$	$T_{m,N}(°C)$	$T_{m,C}(°C)$	$\Delta G(kcal/mol)$
DD*	700	1050	38	65	-3.1
DD-AT*	720	1150	33	50	-2.9
DD-CG*	520	1000	40	40	-3.6
DD**	580	-	-	-	-5.1***
pDD**	560	-	-	-	-8.1***

Table 6.2: Characterization of LC phases for different oligonucleotides forming sticky-end duplexes or blunt-end helices. Data are from ([44], *) and ([99], **), discrepancies on concentration is due to differences in initial salt content: ** data are referred to dialyzed oligomers in MilliQ water. *** simulation results, data extracted from [37].



(a) (b) (c)

Figure 6.4: Polarized microscopy images of LCs domains in DNA-EDC mixtures: (a) 100% N* occupancy at [EDC]/[DNA] = 1; (b) ISO-N* coexistence at [EDC]/[DNA] = 3; (c) Crystals formation at [EDC]/[DNA] = 5

... 5′-CGCGAATTCGCG/CGCGAATTCGCG-3′ ...
... 3′-GCGCTTAAGCGC/GCGCTTAAGCGC-5′ ...

Its structure [41] and collective order into LC phases have been widely characterized [1]. Both 5′-CGCGAATTCGCGp-3′ (DDp) and 5′-pCGCGAATTCGCG-3′ (pDD) have been employed. DD shows LC ordering into N* phase at c=700-1000 mg/ml and ISO-N* transition at $T = 20 - 40°C$, COL phase at c=1000 mg/ml and ISO-COL transition at $T = 40 - 65°C$.

6.4 DNA LCs in presence of EDC

Since LC formation in DNA duplex forming oligomers is a delicate process resulting from a cooperation of end-to-end stacking and mutual alignment of the stacked columns, the presence of other molecular species may well disrupt this process, either weakening the aggregation, or disturbing the orientational alignment of the columns and thus reducing the entropic gain of the system in developing order. Thus it was not obvious if the reagents were compatible with DNA LCs, i.e. whether EDC and HEPES were destabilizing phases formation or not.

HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, CAS Number 7365-45-

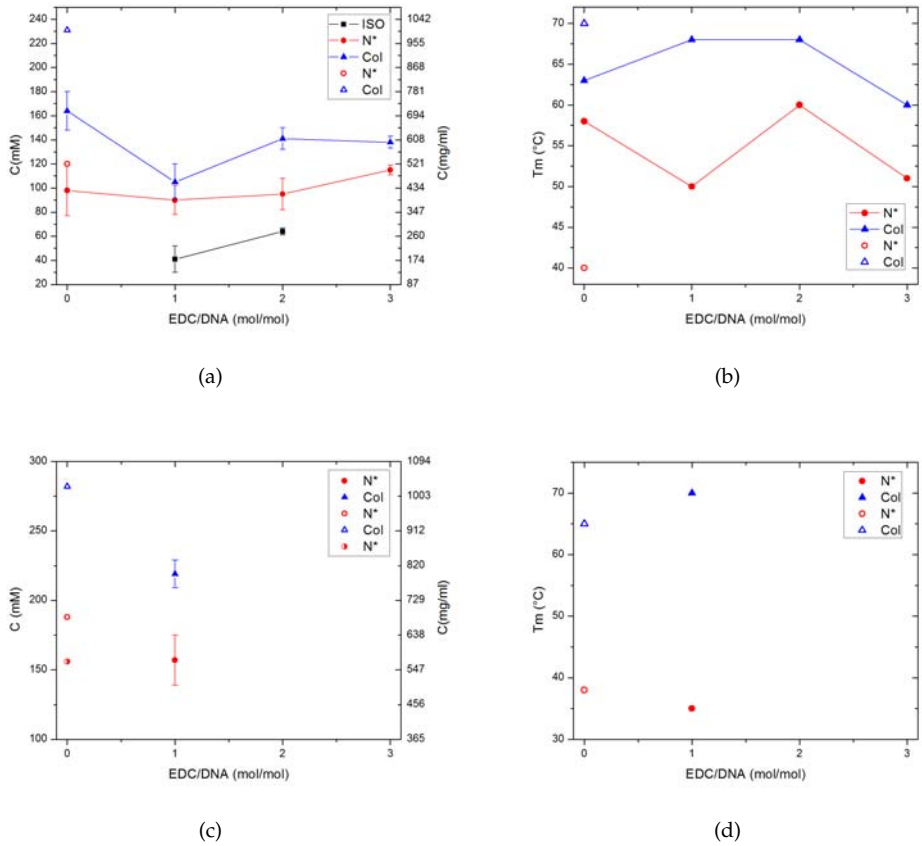


Figure 6.5: Description of the experimental conditions, temperature and concentration, in dependence of the dissolved EDC amount for sticky-ends 14mer $5'-DD-CG-3'$ (a) - (b) and blunt-ends 12mer $5' - DD - 3'$ (c) - (d). Open symbols, from [44], and half-filled symbols, from [99],indicate I*-N and N*-Col phase boundaries at 0 EDC and different ionic strength conditions.

9, from Sigma) buffer is used to keep controlled reaction at pH 7.55. It's commonly used to maintain protein solubility in biochemical experiments and it's an appropriate buffer for these experiments because it has no primary amine groups, that can react with EDC activated phosphates. The desired pH is obtained controlling the amount of sodium hydroxide (NaOH). In accordance with previous experiments of this kind [7], HEPES buffer is used at concentration between 0.01 M and 0.3 M. Extracted Na+ ions concentration is 1.375 mM and 41.25 mM respectively. Salt, even in case of monovalent cations, is known to influence the structure of the DNA helix and to shield electrostatic interaction [46]. As a consequence it influences liquid crystallization. Even if it is still a work in progress, many evidences suggest that salt effects start to be relevant at concentrations similar to the one of the oligomers. At high salt concentration LC phases may also disappear.

EDC (CAS Number 25952-53-8, from Sigma) is the reaction leader, it can form active intermediate with terminal phosphates, so the number of reactive species is closely dependent on the EDC concentration. Thus increasing EDC concentration would increase the possibility to activate a monomer, that would increase the reaction efficiency and the polymerization yield: the more is it possible to activate the chain terminal, the longer could be the final chain fraction. However it is not possible to solubilize in the system any desired amount of EDC. Firstly because EDC solubility in water is less than 100 mg/ml (0.52 M) and secondly Urea is the reaction by-product and it is powerful DNA denaturant, i.e. too much Urea can affect double helices stability and dangling-ends pairing.

In a typical experiment initial low concentration stock solution is prepared with fixed DNA/EDC ratios at constant HEPES. Usually $1 - 2\mu l$ of the solution are evaporated on a microscope glass until LC nucleation starts, than cover slip is added and the system is let equilibrate ($100\% LC$ occupation) and sealed with oil. Final concentration is measured with optical-interferometry methods or by volume measurement. For example, 2.6 mM ($\sim 11 mg/ml$) DD-CGp solution is prepared in 0.01 M HEPES buffer and EDC concentration ranging from 2.6 mM (1x) to 13 mM (5x). For $[EDC]/[DNA] = 2$ N* is found at $410 mg/ml$, i.e. 95 mM, and COL at $610 mg/ml$, i.e. 141 mM; it corresponds to a concentration factor of 37 and 54 times from the initial concentration respectively. Final EDC concentration is 190 mM and 282 mM, HEPES is 0.37 M and 0.54 M.

A systematic study has been performed to determine the maximum amount of EDC still enabling DNA to order into LC phases. Different [EDC]/[DNA] stoichometric ratios have been tested, finding that from 1 to 3 LC phases are maintained. For $[EDC]/[DNA] > 3$ ISO-LC coexistence appears, then if [EDC]/[DNA] exceeds 5 LC formation is inhibited and crystals start to grow (as it is shown in Figure 6.4). In table 6.3 we report reaction conditions adopted in our experiments and those recently used in [7] for comparison.

6.5 Sticky-end ligation

Ligation experiments with sticky-ends 14mer DD-CGp were made both in isotropic and liquid crystal phases. EDC over DNA stoichometric ratio was varied between 1 and 3. In a typical experiment $15\mu l$ of stock solutions at 2.6 mM DNA, EDC (2.6 or 7.8 mM), HEPES 0.01M, are prepared. Then $1\mu l$ is let evaporate on a microscope glass to reach

Reaction Conditions	Sticky LC	Sticky ISO	Blunt ISO - LC	Taran et. al
DNA concentration	90 - 150 mM	2.6 mM	5 - 150 mM	20 mM
Starting DNA	14 mer	14 mer	12 mer	2 - 6 mer
EDC	90 - 450 mM	2.6 - 7.8 mM	5 - 450 mM	0.4 M
HEPES pH 7.55	0.3 M	0.01 M	0.01 M	0.1 M
Temperature	$25°C$	$25°C$	$25°C$	$2°C$
5′ termination	OH	OH	OH	NH_2
[EDC]/[DNA]	1 - 3	1 - 3	1 - 3	20

Table 6.3: Reaction conditions for ligation of sticky-end and blunt-end oligomers in liquid crystal and isotropic phase. Last column reports reaction condition extracted from [7]. LC phases are compatible at least with EDC/DNA ratio equal to 3, setting an upper bound for EDC solubility in these systems.

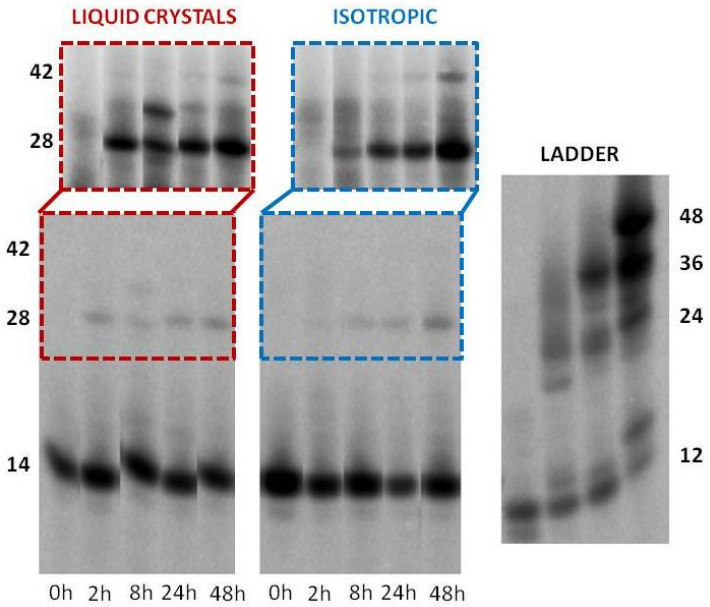
LCs concentration, simultaneously the stock solution is let react. Reaction is stopped by diluting 1 mul of the stock in $1ml$ milliQ water after 4 different intervals of time: 2, 8, 24, 48h. Reaction products have been analyzed by 15% Poly-Acrylamide Gel Electrophoresis in denaturing conditions, as described in Chapter 3. Figure 6.6 shows electrophoretic gels for these samples. Data extraction from gels is obtained by plotting the intensity profile of each gel lanes and measuring the area subtended by each peak, detailed description is reported in Chapter 3.

Both reactions with EDC/DNA ratio 1 and 3 show increase of the ligated product with time and a reaction plateau after 24 hours. This agrees with observation in similar water-soluble carbodiimide condensation reactions reported in literature [7, 95]: the main cause is the fast EDC degradation in aqueous environment due to hydrolysis.

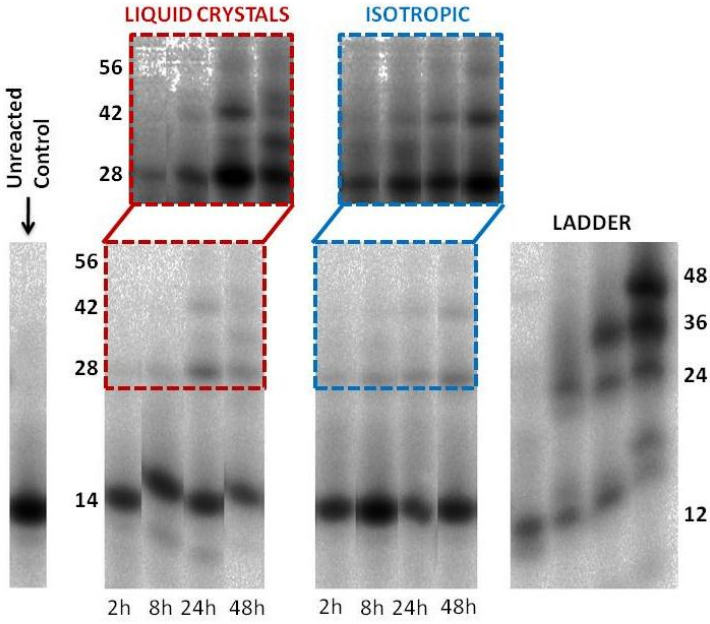
Formation of 56 bases long oligomers, i.e. degree of polymerization equal to 4, is obtained with EDC/DNA ratio equal to 3 after 24 hours. Instead EDC/DNA ratio equal to 1 leads to maximum product length of 42 bases and lower yields. This means that reaction is limited by EDC concentration.

Reaction products are plotted in Figure 6.7 for both ISO (a) and LC (b) in function of products degree of polymerization for EDC/DNA ratio equal to 3. Fitting with the predicted weight fraction distribution from Flory theory of linear polymerization (see Appendix A and Equation A.3) allows to determine the reaction yield p (where $0 < p < 1$), that corresponds to the fraction of the reacted monomers. Another way to estimate the reaction yield is to count the fraction of ligated monomers over the total number of monomers. LC phase shows good reaction yields, 0.30 after 24 hours for N* phase and 0.23 after 48 hours in Col phase. Nevertheless these yields are similar to those obtained in isotropic solutions: 0.23 after 24 hours and 0.28 after 48 hours. At low reaction times, 2 or 8 hours, isotropic phase shows a slightly higher reaction rate, this is mainly due to dimers production, e.g. 16% of the building blocks are condensed into 28 bases long oligomers after 8 hours. Instead LC favors trimers formation reaches 6% after 8 hours versus only 3% in Isotropic phase.

This is the first time that a non-enzymatic ligation reaction is carried on with DNA ordered in liquid crystal phase. Normal protocol conditions adopted in a biochemical re-

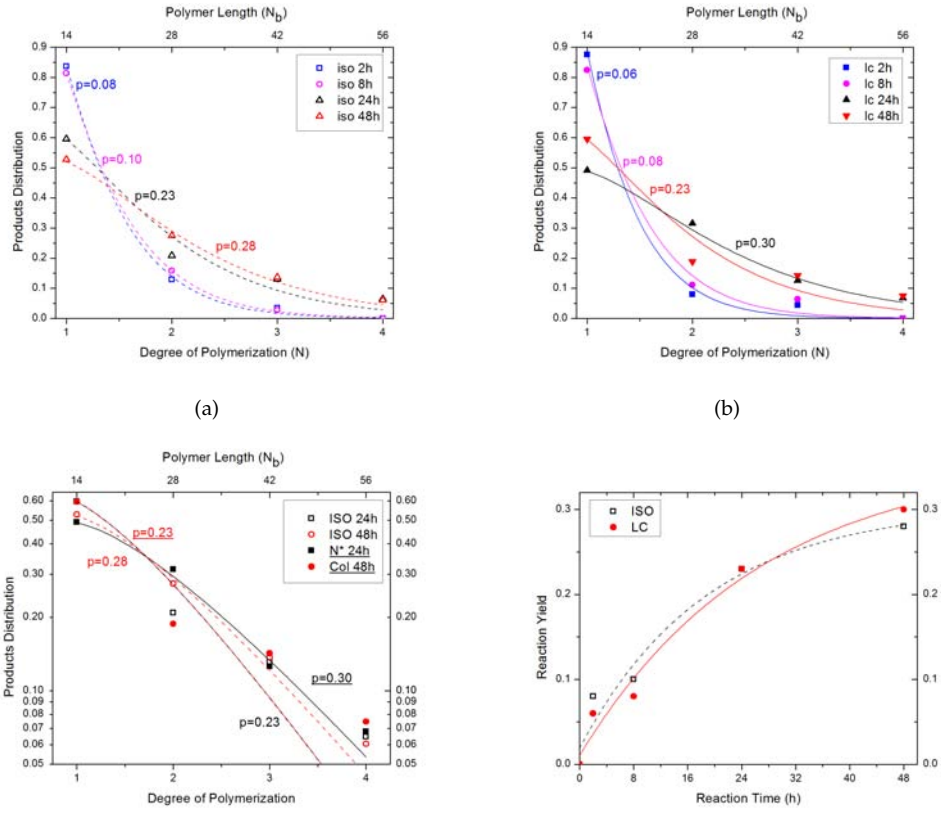


(a)



(b)

Figure 6.6: Denaturing Poly-Acrylamide gel electrophoresis of $\gamma - ^{32} P$ radio labelled products of EDC activated ligation of 5′-DD-CGp-3′. Reactions have been carried on in both Isotropic and LC phase at increasing reaction times: (a) $[EDC]/[DNA] = 1$ and (b) $[EDC]/[DNA] = 3$. Unreacted control sequence is given by incubation of the starting building block without EDC addiction. Ladders are ad-hoc synthesized sequences with repetitions of 12mer DD. Columns have been re-ordered to facilitate reading.



(a) (b)

(c) (d)

Figure 6.7: Products distributions for ligation reactions of sticky DD-CGp at $[EDC]/[DNA] = 3$ as function of aggregate lengths in Isotropic phase (a) and Liquid Crystal phase (b), lines are fits with Flory theory of linear polymerization (see Appendix A); (c) comparison of the products distribution between 24h and 48h reactions in Isotropic and LC phases; (d) plot of reaction yield vs reaction time, shows that after 24h reaction reaches a plateau, lines are a guide for the eyes.

action, usually require dilute oligomers regime (μM or mM), superabundance of reaction activators (enzyme units or chemical activators) and ad-hoc terminal modifications. Although the present experiment was far from these conditions, satisfying reaction yields have been obtained.

It was expected that the higher degree of linear aggregation in LC phase would have brought bigger effect on the products distribution, compared to the more diluted Isotropic regime. Probably this effect is counterbalanced at the same time by the penalty on EDC diffusion, and therefore phosphate activation, in LC more crowded phase. This view is supported by lower dimers production in the first period of the reaction in LCs compared with Isotropic in which EDC can freely diffuse. Instead the yield of longer products (36 mer and 48 mer) slightly increases in LC phase at each reaction timing. Moreover the differences observed in figure 6.7 (c) are confirming the crucial role of EDC mobility: N* phase achieved higher yields than the Isotropic, in 24 hours, instead Col phase has the lowest yields in double reaction time (48 h). Indeed in columnar phase, linear aggregates are subjected to higher degree of orientational and positional order, in which DNA helices are close to be crystallized in a very tight structure, in which EDC diffusion is highly penalized. Therefore, having only orientational order and thus looser packing of DNA helices, N* phase owns the best balance between aggregates length and EDC mobility.

6.6 Blunt-end ligation

The results shown in the previous section suggest that the use of too sticky duplexes would favor aggregates formation even in Isotropic phase. It has been recently shown that for linear aggregation of bifunctional patchy particles, the increase of the average chain length upon transition from Isotropic to Nematic is more effective at low end-to-end adhesion of the building blocks [46]. Thus, having lower stacking interaction, blunt-end DNA helices are good candidates in order to enlarge the gap in the aggregate length between Isotropic and LC phases. For this reason ligation experiments with blunt-end 12mer DD were made both in isotropic and liquid crystal phases.

Three different experiments have been carried on involving blunt-end:

1. **Concentration scale**

   Starting from diluted regime, a series of experiments were performed with increasing DNA amount, at constant EDC/DNA ratio, in order to reach high concentration phases. This allow to gradually sample the length distribution of the aggregates both in Isotropic and in approaching ISO-LC transition and in different LCs phases. EDC over DNA stoichometric ratio was varied between 1 and 3, higher ratios destabilize LCs (see section 6.4).

2. **Mixtures of single and double strands** In order to overcome the strict limitation on EDC amount that can be added in LC phases, there is the need of systems in which an EDC reservoir is placed in equilibrium with LC phases without destabilizing them. A proper way is represented by mixture of complementary oligomers,

forming double helices, and unpaired single strands. In such system the complementary oligomers tend to be incorporated into the LC domains, and segregated from the surrounding isotropic phase of non-complementary strands [2]. Thus larger amount of EDC can be dispersed in single strand rich Isotropic phase. Nevertheless the peculiar fact that the constituent of both phases belong to the same molecular species has two linked implications:

- single strands are not only deplectants but also competitors in ligation reaction, thus can interfere with the duplexes polymerization;
- it can be a suitable system to test the emergence of selective linkages trough the physical segregation.

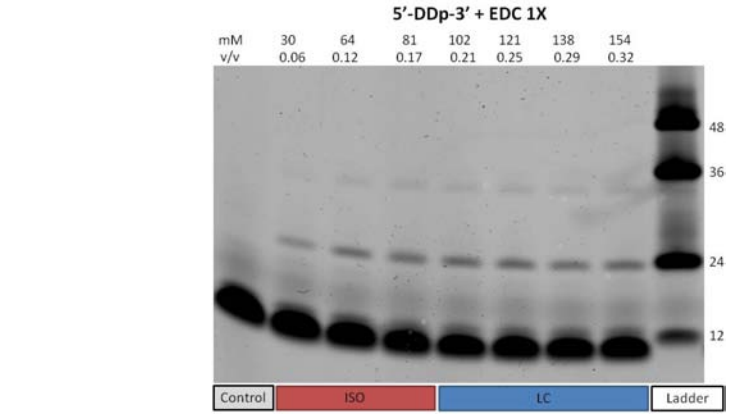
3. **PEG induced phase separation** Another way to increase EDC amount is to exploit the phase separation that occurs when DNA is mixed with Poly-(ethylene glycol) (PEG) [2]. The entropic gain given by the separation of species differing in shape or flexibility, induces segregation of a DNA rich phase, that can order into LCs, from a PEG rich Isotropic phase. This seems a very convenient way to approach the problem because:

   - the concentration of DNA can be controlled by the concentration of PEG. This is a well-known effect since the concentration of PEG determines the osmotic pressure of the solution [100, 101];
   - coexisting isotropic PEG phase constitutes a suitable container for large amount of EDC, that is free to diffuse and to reach the neighboring DNA duplexes rich phases. In addiction reaction by-products as urea can follow the inverse path and disperse into PEG phase, reducing the destabilizing effect on DNA helices;
   - PEG doesn't interfere with EDC ligation reaction that involves exclusively DNA duplexes.

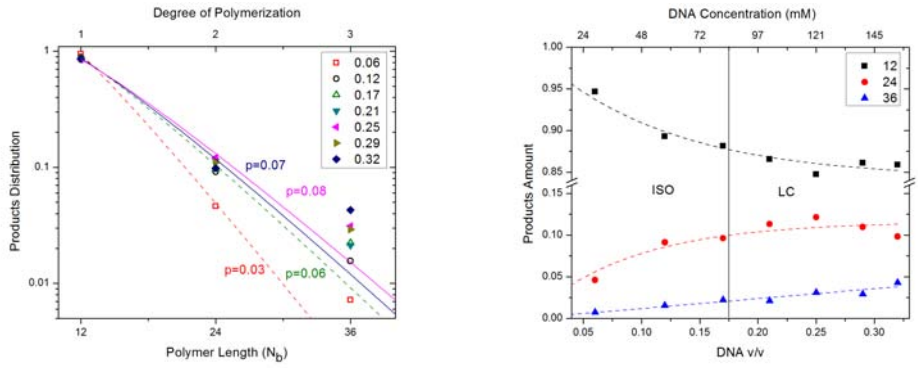
6.6.1 Concentration scale

Ligation experiments were made varying building blocks concentration in the range $10 - 600 mg/ml$, i.e. 3 - 154 mM, i.e. 0.006 - 0.32 volume fraction (DNA volume/total solution volume). From now on we refer only to DNA molarity or volume fraction, for simplicity of comparison with results of next paragraphs. Comparison between phosphate termination at 5′ (pDD) and at 3′ (DDp) was made to find the best reaction conditions.

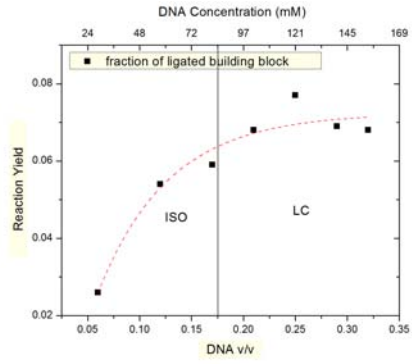
Reactions were performed in capillary or in eppendorf tubes, in a typical experiment the right amount of lyophilized oligomers is hydrated with freshly prepared EDC solution in pH 7.55 HEPES buffer. Samples are sealed by siliconic oil to prevent evaporation. Extra carbodiimide addiction has been achieved in some cases by EDC dispersion in siliconic oil: EDC is not soluble in oil but it has been observed to sediment, reaching water-oil interface. Reaction is stopped by 20 fold dilution in 50mM Ethanolamine after



(a)

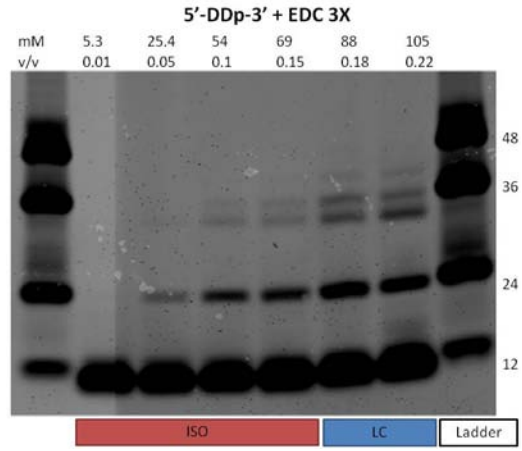


(b) (c)

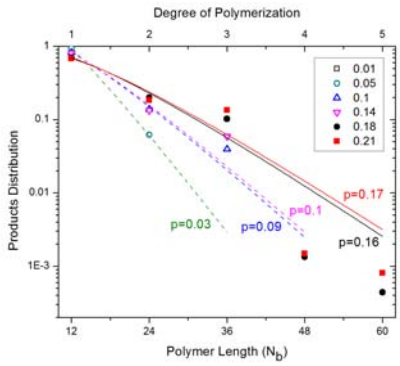


(d)

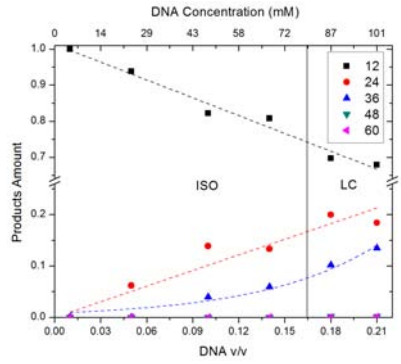
Figure 6.8: (a) Denaturing Poly-Acrylamide gel electrophoresis of ligation products obtained from 5'-DDp-3' building blocks at $[EDC]/[DNA] = 1$. See reaction conditions in 6.4. (b) Distribution of the polymerization products vs length, open symbols are from ISO, filled from LC, lines are fits are with Flory theory of linear polymerization (see Appendix A); (c) Plot of products amount vs DDp volume fraction, lines are a guide for the eyes; (d) Reaction yield vs DDp volume fraction shows a plateau at high concentrations.



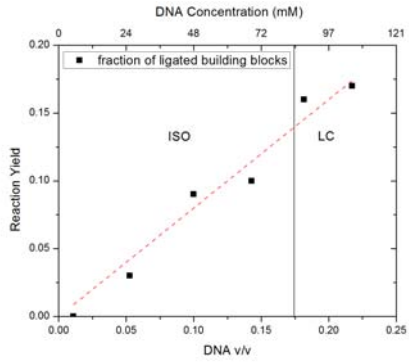
(a)



(b)



(c)

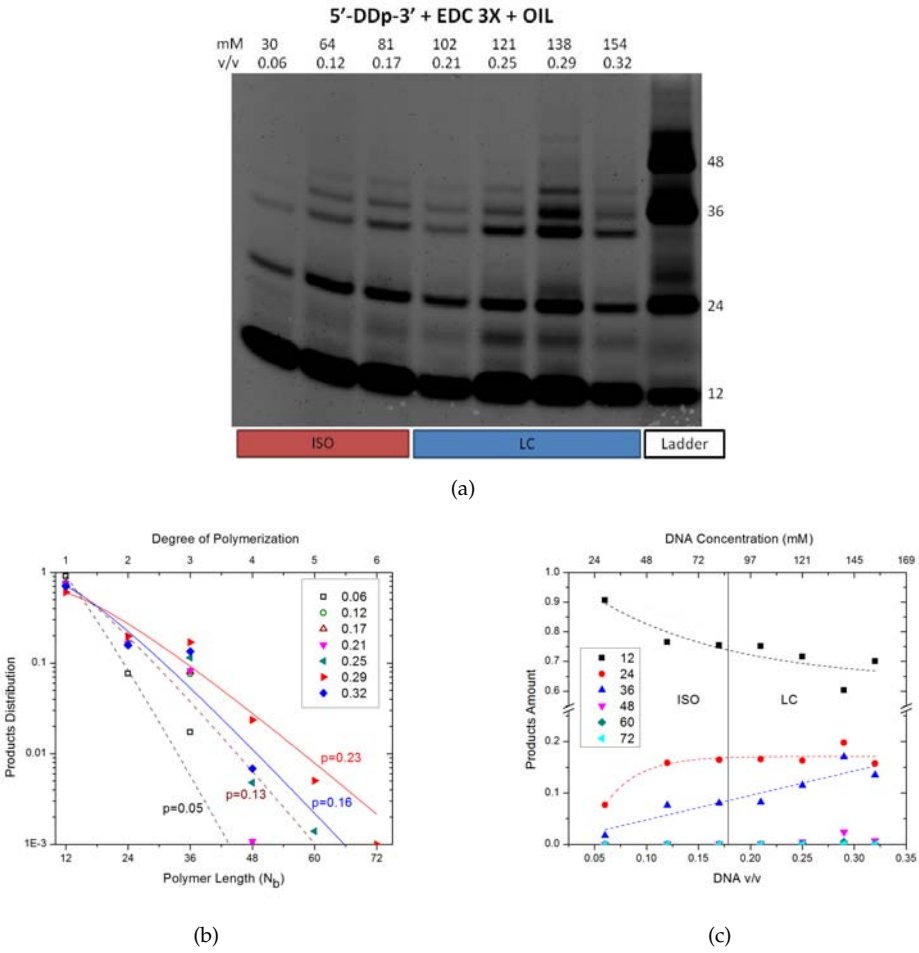


(d)

Figure 6.9: (a) Denaturing Poly-Acrylamide gel electrophoresis of ligation products obtained from 5'-DDp-3' building blocks at $[EDC]/[DNA] = 3$. See reaction conditions in 6.4. (b) Distribution of the polymerization products vs length, open symbols are from ISO, filled from LC, lines are fits with Flory theory of linear polymerization (see Appendix A); (c) Plot of products amount vs DDp volume fraction, lines are a guide for the eyes; (d) Reaction yield vs DDp volume fraction, lines are a guide for the eyes.

Sample	DDp 1x	DDp 3x	DDp 3x + oil	pDD 3x
DNA concentration	30 - 154 mM	5.3 - 105 mM	30 - 154 mM	30 - 154 mM
EDC concentration	90 - 462 mM	16 - 315 mM	90 - 462 mM	90 - 462 mM
[EDC]/[DNA]	1	3	3 (+ 5 in oil)	3
Terminal Phosphate	3′	3′	3′	5′
Reaction Time (h)	24	24	24	24
Volume (μl)	1	1	1	1

Table 6.4: Reaction conditions for ligation of self-complementary 12 mer 5′-CGCGAATTCGCG-3′ at increasing building blocks concentration.

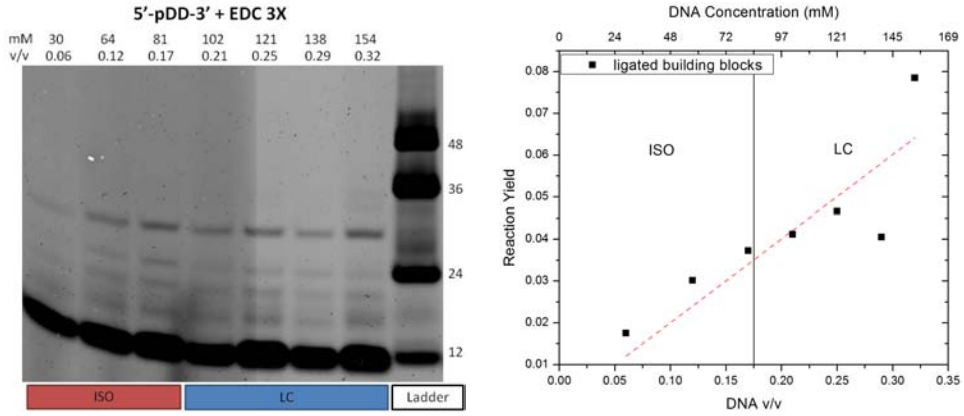


(a)

(b)

(c)

Figure 6.10: (a) Denaturing Poly-Acrylamide gel electrophoresis of ligation products obtained from 5′-DDp-3′ building blocks at $[EDC]/[DNA] = 3$ and EDC addiction in siliconic oil. See reaction conditions in 6.4. (b) Distribution of the polymerization products vs length, open symbols are from ISO, filled from LC, lines are fits with Flory theory of linear polymerization (see Appendix A); (c) Plot of products amount vs DDp volume fraction, lines are a guide for the eyes.



(a) (b)

Figure 6.11: (a) Denaturing Poly-Acrylamide gel electrophoresis of ligation products obtained from 5'-pDD-3' building blocks at $[EDC]/[DNA] = 3$. See reaction conditions in 6.4. (b) Reaction yield vs pDD volume fraction.

24 hours reaction time. Reaction products have been analyzed by 15% Poly-Acrylamide Gel Electrophoresis in denaturing conditions. Reaction details are reported in table 6.4.

At EDC/DNA = 1 (Figure 6.8) ligation happens with low efficiency, maximum yield is 0.08, products distribution is exponential independently on initial concentration. Reaction yield follows a decaying exponential, this can be caused by the fact that reaction is limited by low EDC amount: EDC degradation during the reaction time could inhibit the formation of a relevant amount of longer strands.

Raising of EDC/DNA ratio up to 3 (Figure 6.9), the reaction becomes more efficient, ligation yield reaches 0.17, and it is less affected by EDC consumption on the experiment time scale. Even if reaction yield linearly scales with concentration it's possible to recognize some effect of the LC formation in the broadening of products distribution and the formation of longer polymers, as 48 mer and 60 mer strands (Figure 6.12).

In order to maximize this contribution, experiments with EDC addiction in oil have been performed (Figure 6.10). Increasing the reaction starter amount brings to the formation of longer products, 60 mer and 72 mer strands are found only at LC concentration, even if with very small yields. The highest reaction yield is 0.23 and it is reached in LC phase at DNA volume fraction 0.29.

The use of 5'-phosphorilated oligomers (Figure 6.11) has always brought to lower reaction yields confirming the lower reactivity of this configuration, as already reported by [94]. For this reason we opted to use only 3'-phosphate Dickerson Dodecamers for next experiment.

The performed experiments with increasing building block concentration show weak evidences of LC phase contribution to reaction yields: the most relevant findings are that formation of longer strands is achieved in LCs (see Figure 6.12b) and consequently broader aggregates distribution is observed. The causes of this behavior could be found

considering two competing phenomena related to the high close packing of oligomers in LC phase: on one hand increase in degree of aggregation of linear physical aggregates helps in keeping the terminals to be ligated in proximity, on the other hand increasing of the local viscosity penalizes EDC diffusion, affecting the reaction initiation.

The comparison of reaction yields at different [EDC]/[DNA] stoichiometric ratio confirms this view (figure 6.12a). The fraction of ligated building blocks has been plotted vs DNA v/v. Comparing the slopes of the linear fit for [EDC]/[DNA]=1 and [EDC]/[DNA]=3 it is possible to infer:

- $m_1 = 0.28$ and $m_3 = 0.69$: the slope of linear fit on the whole data range increases with [EDC]/[DNA], i.e. the efficiency of the reaction depends on EDC ammount, as it was expected;
- $m_1^{ISO} = 0.39$, $m_3^{ISO} = 0.83$, $m_1^{LC} = 0.05$, $m_3^{LC} = 0.34$: the slope of the linear fit in LC is always lower than that in ISO, i.e. something, such as EDC diffusion, is limiting the growth of ligation rates in LCs;
- the ratios of the slopes in ISO, $r_{ISO} = m_3^{ISO}/m_1^{ISO} = 2.1$, is lower than that in LC, $r_{LC} = m_3^{LC}/m_1^{LC} = 6.4$: thus increasing EDC amount by a factor of 3, the growth of the reaction yield in LC phase shows an increase of a factor r_{LC} that is bigger than 3, instead the increase in ISO is of a factor r_{ISO} lower than 3.

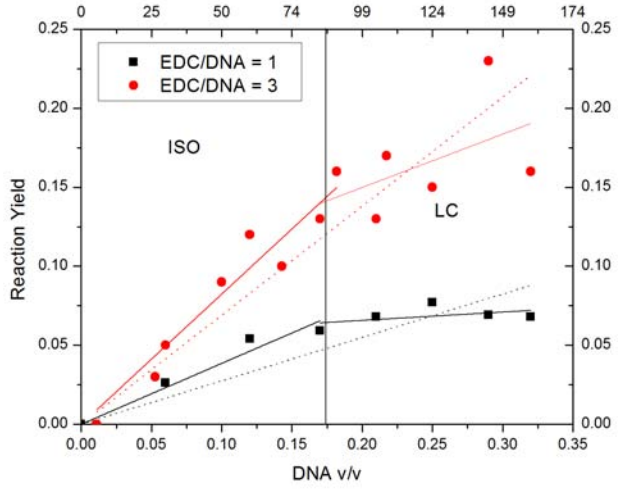
Therefore these findings support the view that the limiting factor for ligation in ISO is not the monomers activation (ruled by EDC amount) but rather the high distance between the monomers to be ligated, thus enhancing the probability of EDC activated phosphate to be subjected to hydrolysis before meeting another monomer. On the other hand the fact that LC phase shows the opposite tendency supports the fact that the limiting factor here is linked to the process of monomers activation by EDC.

6.6.2 Mixtures of single and double strands

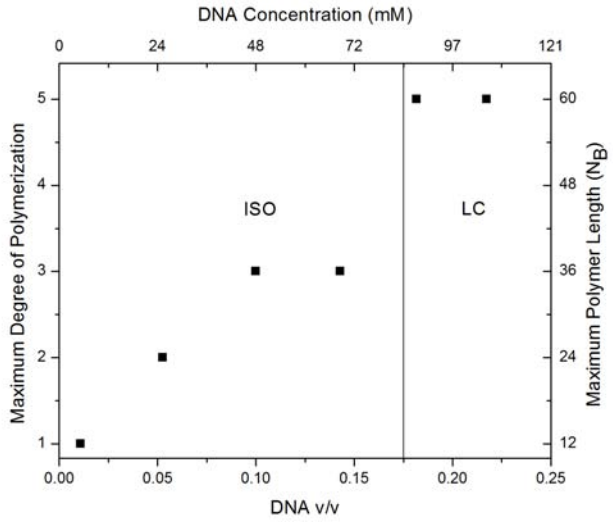
Mixtures of single and double strands are obtained by the addiction of 16 mer poly-T DNA strands, 5'-TTTTTTTTTTTTTTTT p-3' (16Tp), in solution with 3' phosphate Dickerson Dodecamer (DDp). The lack of the complementary poly-A strand causes 16Tp to stay in single strand conformation while self-complementary DDp is always able to hybridize forming double helix.

Three different DDp concentrations have been tested, 16 mM, 32 mM and 40 mM, with increasing 16Tp concentration from 9 mM to 90 mM in order to keep fixed stoichometric ratios between single strands and double helices, respectively 1, 2, 3 and 4. EDC concentration is chosen to maintain constant [EDC]/[DDp] = 20. Nevertheless both DDp and 16Tp can undergo EDC activation, thus single strands are not only deplectants but also competitors in ligation reaction. At low DDp and 16Tp concentrations no LCs are observed and the system is fully isotropic, instead at increasing DDp and 16Tp concentrations LCs start to form and the system develops phase separation.

In figure 6.13 gel electrophoresis of the reacted samples is reported. Analysis of the gel bands shows a peculiar behavior: in isotropic samples the production of 36 mer aggregates is higher than that of 24 mer. For a linear condensation polymerization reaction

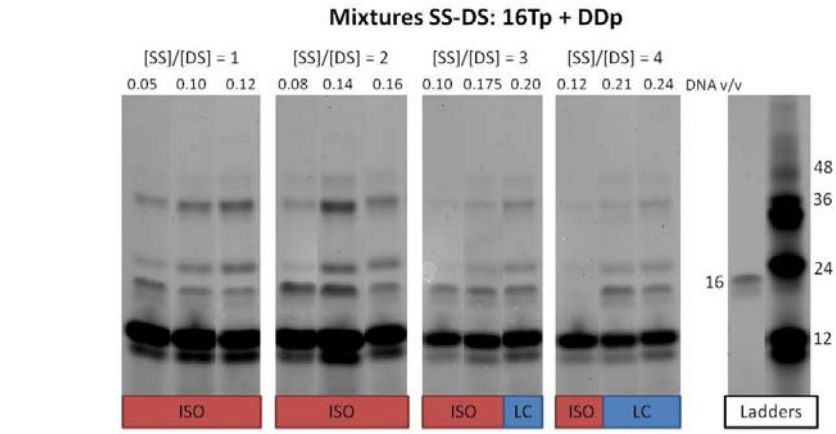


(a)

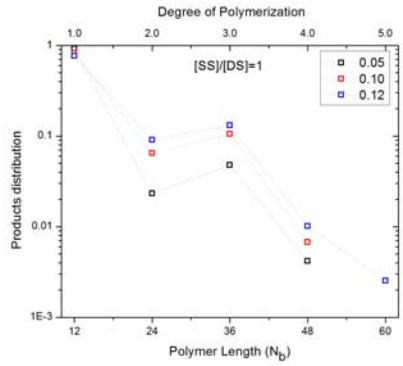


(b)

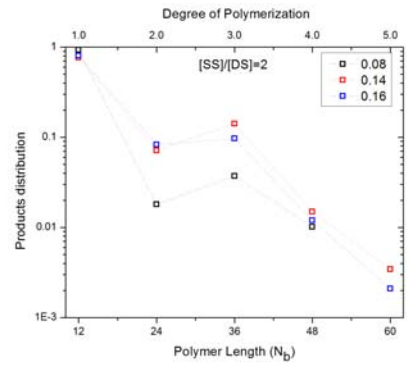
Figure 6.12: Results of concentration scales. (a) Comparison of reaction yields at different [EDC]/[DNA] with linear fit of the whole range of data (dotted) and divided in ISO and LC ranges (solid). It is evident the decrease of the fit slope in LC region in comparison with ISO. This can be due to inefficiency of EDC diffusion in highly packed LC environment. (b) Plot of the maximum polymer length vs DDp volume fraction for ligation products obtained from 5′-DDp-3′ building blocks at $[EDC]/[DNA] = 3$. The appearance of the longest polymers is achieved at LC phase formation.



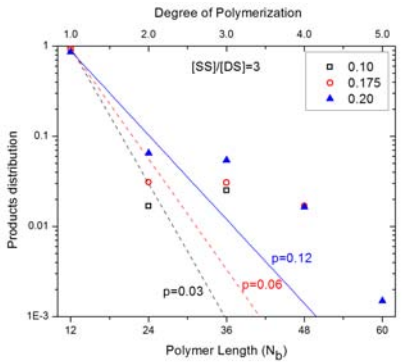
(a)



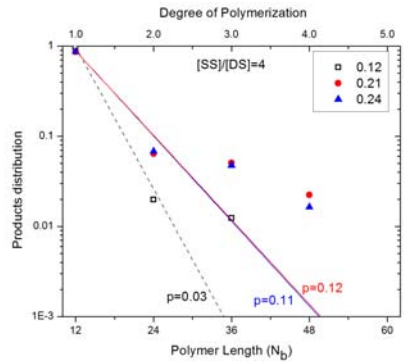
(b)



(c)



(d)



(e)

Figure 6.13: (a) Poly-Acrylamide gel electrophoresis of ligation products obtained from 5′-DDp-3′ building blocks mixed with 5′-16Tp-3′ single strands at increasing stoichometric ratios (rows have been reordered to facilitate reading); (b) & (c) Distribution of the polymerization products vs length, for ISO samples: unexpected non-monotonic yield of 36 mer aggregates is noticeable; (d) & (e) Distribution of the polymerization products vs length, open symbols are from ISO, filled symbols from LC, normal monotonic behavior is restored. Lines are fits with Flory theory of linear polymerization (see Appendix A).

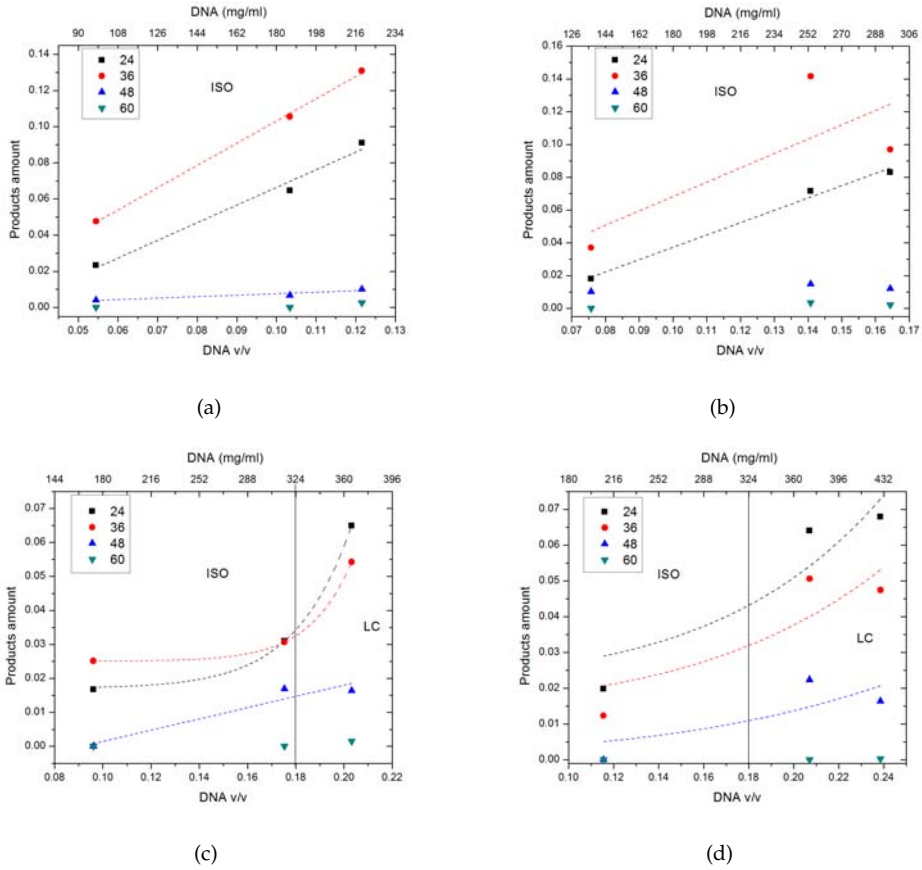


Figure 6.14: Plot of products amount vs total DNA volume fraction for 5′-DDp-3′ and 5′-16Tp-3′ mixtures. (a) & (b) yield of 36 mer aggregates is higher than 24 mer in ISO phase; (c) & (d) yield of 36 mer aggregates becomes lower than 24 mer at ISO-LC transition.

it is expected a monotonic decreasing distribution of the products [102], the emergence of a peak at a certain length, higher than the monomer, can happen only under consumption of the starting building block during the reaction time, but this is not the case because monomers are still present in big quantity.

Observing electophoretic mobility of poly-T 16 mer, that is very close to 24 mer mobility, we can assume that strands formed by DDp and 16Tp can have mobility similar to the 36 mer strand, leading to superposition of the gel bands. In LC samples normal behavior is restored, suggesting that the rise of phase separation favors selectivity of ligation reaction: ligation between stacked helices is more probable than ligation of double helices with single strands. Future experiments involving selective enzymatic digestion of poly-T strands and subsequent run of the remaining material would help testing this hypothesis.

6.6.3 PEG induced phase separation

Mixtures of DDp and PEG polymers have been studied. Ligation experiments were made at fixed building blocks concentration and varying PEG volume fraction. High PEG volume fraction leads to phase separation between PEG rich phase and LC phase rich in DNA. However at low PEG contents there are no evidences of ISO-ISO phase separation between DNA phase in coexistence with PEG rich phase. In these systems phase separation is apparently correlated with the emergence liquid crystal order.

Reactions were performed in capillary or in eppendorf tubes. In a typical experiment the right amount of lyophilized oligomers is hydrated with PEG solution containing fresh prepared EDC solution in pH 7.55 HEPES buffer. PEG polymers of different molecular weight have been tested. Samples are sealed by siliconic oil to prevent evaporation. Reaction is stopped by 20 fold dilution in 50mM Ethanolamine after 24 hours reaction time. Reaction products have been analyzed by 15% Poly-Acrylamide Gel Electrophoresis in denaturing conditions. Reaction details are reported in table 6.5.

Reaction Conditions	EDC 10x	EDC 100x	PEG 1K
DNA concentration	3 mM	6 mM	1.5 - 2.7 mM
EDC concentration	30 mM	600 mM	15 - 27 mM
[EDC]/[DNA]	10	100	10
PEG MW (Dalton)	8000	8000	1000
Reaction Time (h)	24	24	24
Volume (mul)	2	5	15

Table 6.5: Reaction conditions for mixtures of self-complementary 12 mer 5′-CGCGAATTCGCGp-3′ and PEG.

Impressive results have been obtained with PEG 8000 at increasing volume fraction from 0.08 to 0.39. As shown in Figure 6.15, LC formation is correlated to huge increase in ligation yields. If at low PEG volume fractions only dimers formation is observed, longer ligated products appear as soon as the system segregates forming LC domains. Maximum aggregates length shows a step like behavior from 24 to 84 bases, i.e. 7 times the starting building block length, at the ISO-LC transition, as shown in Figure 6.18. At the same time products distribution is subjected to a wide broadening. Both the decrease of the monomers amount and the increase of longer aggregates amount follow exponential laws vs PEG volume fraction, (see Figure 6.15). Reaction yield has an exponential increases at LC formation rising from ~ 0.06 to 0.38, as shown in Figure 6.18.

Increasing EDC/DNA stoichometric ratio to 100, figure 6.16, polymerization of up to 12 monomers forming a 144 bases long strand has been achieved in LC phase. In this system monomers conversion is close to 0.50 leading to the shift of the distribution to higher degree of polymerization.

Lower molecular weight PEG (1000 Dalton) is showing similar effects on reaction: LCs formation and segregation at high PEG volume fractions causes the increase of products length and broadening of the products distribution (figure 6.17).

These results highlight the crucial role of compartmentation for ligation reactions

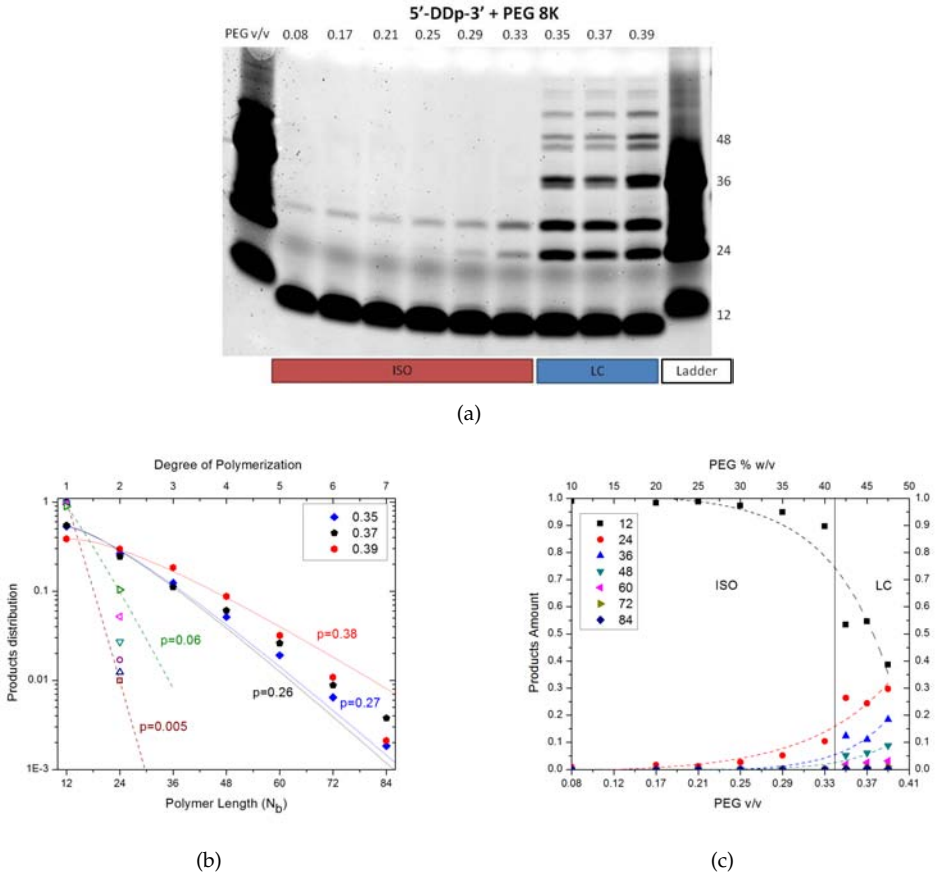
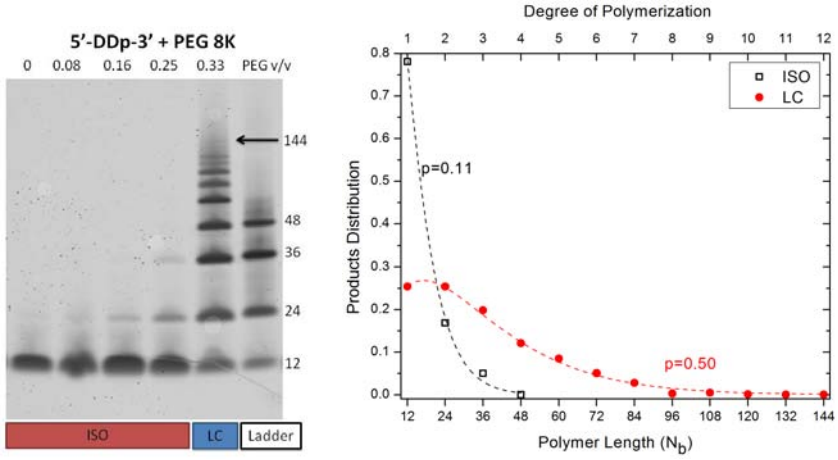


Figure 6.15: (a) Denaturing Poly-Acrylamide gel electrophoresis of ligation products obtained from 5′-DDp-3′ building blocks mixed with PEg 8000. See reaction conditions in 6.5. (b) Distribution of the polymerization products vs length, open symbols are from ISO, filled from LC, lines are fits with Flory theory of linear polymerization (see Appendix A); (c) Plot of products amount vs PEG volume fraction, lines highlight the exponential growth of ligation products longer than 12 bases, while monomers are decaying.

involving short DNA helices. In mixtures of DNA oligonucleotides and PEG, the emergence of liquid crystal order leads the segregation of highly concentrated DNA phase from the rest of the environment. Such increase in building blocks local concentration and the proximity of terminal groups of stacked duplexes are the two mechanisms that mainly lead to increase of reaction efficiency.

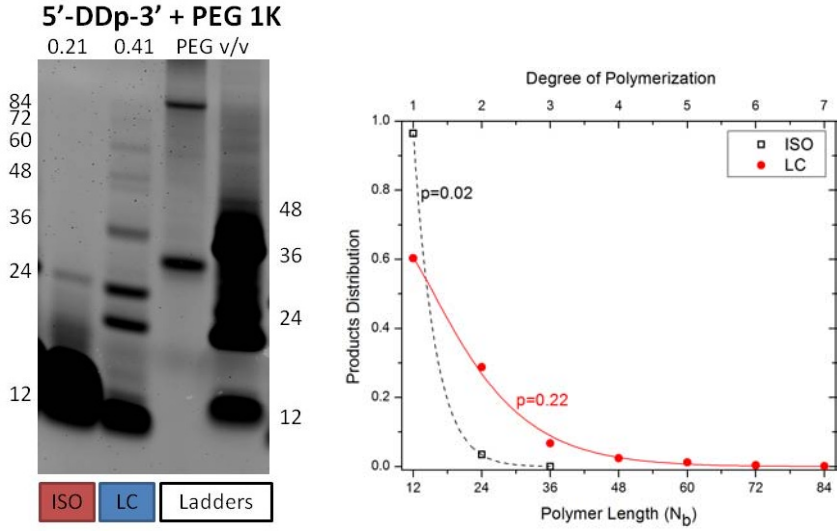
6.7 Conclusions

Data here reported demonstrate that EDC-driven ligation of short complementary DNA strands has been achieved (and in some cases also largely catalyzed) in liquid crystalline phase. The use of water-soluble carbodiimides as condensing agent is widely diffused



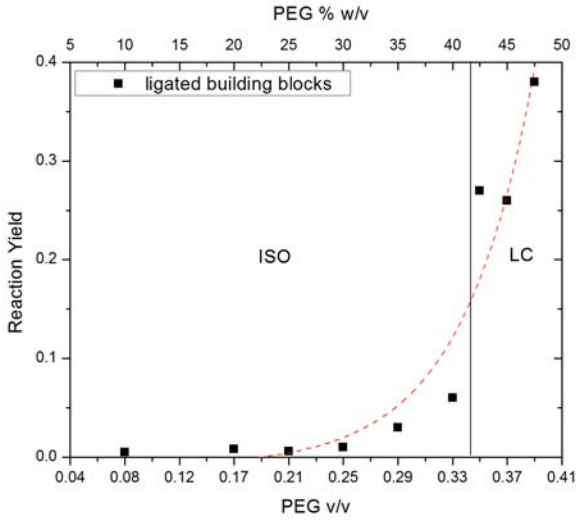
(a) (b)

Figure 6.16: (a) Denaturing Poly-Acrylamide gel electrophoresis of ligation products obtained from 5'-DDp-3' building blocks mixed with PEg 8000. See reaction conditions in 6.5. (b) Distribution of the polymerization products vs length, open symbols are from ISO, filled from LC, fit are with Flory theory of linear polymerization (see Appendix A).

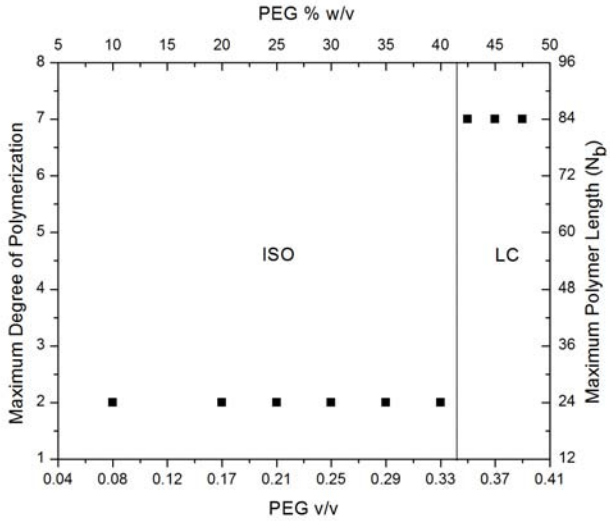


(a) (b)

Figure 6.17: (a) Denaturing Poly-Acrylamide gel electrophoresis of ligation products obtained from 5'-DDp-3' building blocks mixed with PEg 1000. See reaction conditions in 6.5. (b) Distribution of the polymerization products vs length, open symbols are from ISO, filled symbols from LC, fit are with Flory theory of linear polymerization (see Appendix A).



(a)



(b)

Figure 6.18: Impressive results obtained from 5′-DDp-3′ building blocks mixed with PEg 8000 that clearly show how ligation products increase both quantitatively and qualitatively at LC formation: (a) Reaction yield vs PEG volume fraction has a pronounced exponential growth that correlate with LC formation; (b) Maximum aggregates length vs PEG volume fraction shows a clear step behavior at ISO-LC transition.

nowadays for oligonucleotides linkage, but LCs are a completely new environment for this kind of reaction.

Despite having adopted the choice of not introducing modified molecular terminals, such as the highly reactive amino-termination, and despite strict constraints on EDC amount, good reaction yields have been found for the polymerization of DNA oligomers both in isotropic than in LC phases.

Comparison of sticky-end and blunt-end building blocks confirmed the crucial role of end-to-end adhesion: higher interaction energy favor higher reaction yields, as shown by sticky-end data. On the other hand low interaction energy allows to make the effect of ISO-LC transition critical in favoring the production of longer polymers.

The performed experiments with increasing building block concentration show weak evidences of LC phase contribution to reaction efficiency (see Figure 6.12a), nevertheless the formation of longer strands is found prevalently in LCs (see Figure 6.12b), and consequently broader aggregates distribution is achieved. The causes of this behavior is found to be due to two competing phenomena related to the high close packing of oligomers in LC phase: on one hand increase in degree of aggregation of linear physical aggregates help in keeping the terminals to be ligated in proximity, on the other hand increasing of the local packing penalizes EDC diffusion, affecting the reaction starting point.

Ligation achieved in mixtures of single and double strands suggests a promising path for testing the arising of selectivity mediated by phase separation that need more investigation in the future.

The most remarkable effects are found in mixtures of DNA oligomers with PEG: LC nucleation causes higher reaction yields, up to 50% monomer conversion, and a sharp increase in the degree of polymerization (see Figure 6.18), up to lengths of 144 bases, i.e. 12 times the initial length (see Figure 6.16). These evidences confirm the crucial role of self-assembly of short DNA duplexes into LC phases in promoting the segregation of highly concentrated DNA phase in a molecular mixture. Such increase in building blocks local concentration, in a system that allow adequate presence of reaction initiators, leads to large increase of reaction efficiency.

APPENDIX A

Molecular Size Distribution in Linear Condensation Polymers

This appendix report Flory theory on molecular size distribution in linear condensation polymers. The results of this simple theory have been used to fit the products distribution of the ligation reaction described before in the text.

The theory of the size distribution of linear polycondensation polymers was formulated by P. J. Flory in works published from 1936 to 1948. This theory offers a theoretical analysis of the molecular size distribution in linear condensation polymers formed by the inter-molecular reaction of bifunctional compounds [102].

Here I report only about polycondensation with an equivalent ratio of functional groups, i.e. the case that corresponds to ligation reaction of double strand DNA oligomers via 3′-5′ phosphodiester linkage. Infect each DNA duplexes owns both reacting groups: Phosphate group and hydroxyl group.

Assumption of Flory theory

Flory analyzed polymerization reactions that involve the coupling of two bifunctional compounds to form a bifunctional compound of higher molecular weight, which in turn is capable of polymerizing with other molecules. Assuming that the reactions: monomer with monomer, monomer with polymer, and polymer with polymer all involve the same chemical process, and with the fundamental assumption that the reactivity of a functional group is independent of the size of the molecule to which it is attached. It is assumed that new bonds arising during the course of the reaction have the same resistance to the reverse reaction. It is also assumed that the formation of cyclic compounds (as well as other side reactions) does not occur to an appreciable extent. The errors arising from the assumptions will in general be most significant in the range of low molecular weights; for polymers of high average molecular weight, where the distribution is broad and the low molecular weight portion is small, deviation from the calculated results should not be large.

Definitions

A reactant building block molecule is defined as a "segment" and the term x-mer will denote a polymer composed of x monomers. N_0 is the total number of segments. N is the number of unreacted monomers at the time at which reaction is stopped. Thus $N_0 - N$ is the number of reacted monomers. p is the extent of reaction, i.e. the fraction of the total number of reacted monomers $p = (N_0 - N)/N_0$ (in this thesis it is assumed as synonymous of reaction yield). From this follow that $N_0 = N(1-p)^{-1}$.

Fundamental distribution functions

p can be viewed as the probability that any monomer can react during the reaction time, $1 - p$ is instead the probability that the monomer doesn't react.

Thus the formation of a x-mer needs the realization of $(x–1)$ bonds, i.e. $(x-1)$ times probability p, and two non bonds, i.e. two times probability $(1-p)$:

$$\pi_x = p^{(x-1)}(1-p)^2 \tag{A.1}$$

The number of methods that enable the formation of an x-mer is equal to the number of combinations of x over $(x-1)$:

$$C_x^{x-1} = \frac{x!}{(x-1)![x-(x-1)]!} = x \tag{A.2}$$

Thus follow that the probability for the formation of any of the x configurations is :

$$\Pi_x = xp^{(x-1)}(1-p)^2 \tag{A.3}$$

where Π_x is the weight fraction distributions and is defined as the total number of monomers which exist as components of x-mers over the total number of monomers (N_0).

Therefore the number of x-mer polymer is given by:

$$N_x = \frac{N_0 \Pi_x}{x} = N_0 p^{(x-1)}(1-p)^2 \tag{A.4}$$

from which, defining P_x as the mole fraction distribution of x-mers, it results:

$$P_x = N_x/N = p^{(x-1)}(1-p) \tag{A.5}$$

Differentiating P_x, with respect to x and equating to zero

$$\frac{\partial P_x}{\partial x} = p^{(x-1)}(1-p)\ln p = 0 \tag{A.6}$$

Since $x = \infty$, is the only solution, P_x is monotonic.

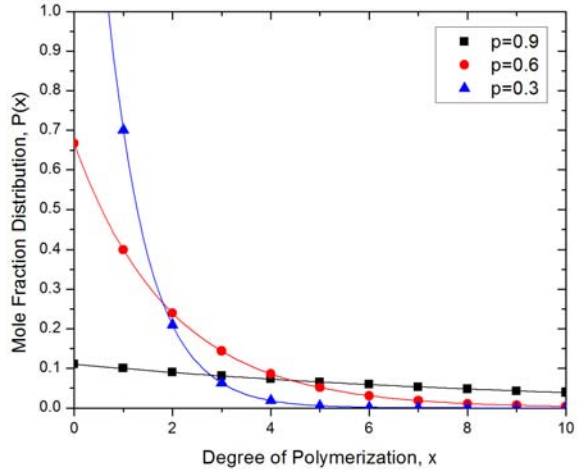


Figure A.1: Mole fraction distributions (Equation A.6) for different value of the parameter p

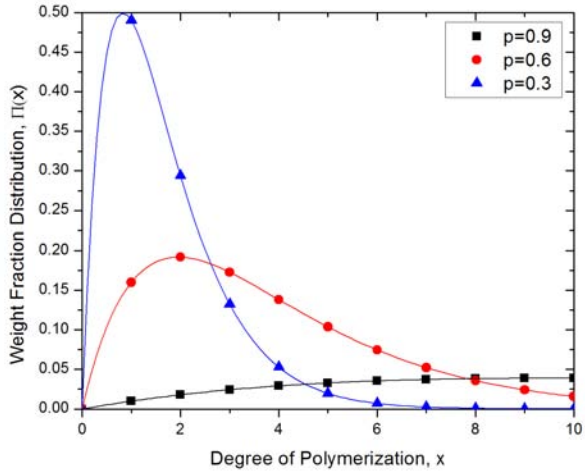


Figure A.2: Weight fraction distributions (Equation A.3) for different value of the parameter p

Bibliography

[1] M. Nakata, G. Zanchetta, T. Bellini, and N. A. Clark, *End-to-End Stacking and Liquid Crystal Condensation of 6 to 20-Base Pair DNA Duplexes.*, Science **318** (2007) 1276.

[2] G. Zanchetta, M. Nakata, M. Buscaglia, T. Bellini, and N. A. Clark, *Phase separation and liquid crystallization of complementary sequences in mixtures of nanoDNA oligomers.*, Proc. Natl. Acad. Sci. U.S.A. **105** (2008) 1111.

[3] S. L. Miller, *A production of amino acids under possible primitive earth conditions.*, Science **117** (1953) 528.

[4] A. P. Johnson, H. J. Cleaves, J. P. Dworkin, D. P. Glavin, A. Lazcano, and J. L. Bada, *The Miller volcanic spark discharge experiment.*, Science **322** (2008) 404.

[5] G. F. Joyce, *The antiquity of RNA-based evolution.*, Nature **418** (2002) 214.

[6] C. De Duve, *Singularities - Landmarks on the pathways of life.* Cambr. Univ. Press., 2005.

[7] O. Taran, O. Thoennessen, K. Achilles, and G. Von Kiedrowski, *Synthesis of information-carrying polymers of mixed sequences from double stranded short deoxynucleotides.*, Jour. Sys. Chem. **1** (2010) 9.

[8] C. R. Calladine, H. Drew, B. Luisi, and A. Travers, *Understanding DNA: the molecule and how it works.* Elsevier Academic Press, New York, 2004.

[9] T. Bellini, R. Cerbino, and G. Zanchetta, *DNA-based soft phases*, Top Curr Chem **318** (2012) 225.

[10] R. Podgornik, H. H. Strey, and V. A. Parsegian, *Colloidal DNA*, Curr Opin Colloid Interface Sci **3** (1998) 534.

[11] J. Santalucia Jr., *A unified view of polymer, dunbell, and oligonucleotide DNA nearest-neighbor thermodinamics.*, Proc. Natl. Acad. Sci. USA **95** (1998) 1460.

[12] J. Santalucia Jr. and D. Hicks, *The Thermodinamics of DNA Structural Motifs.*, Annu. Rev. Biophys. Biomol. Struct. **33** (2004) 415.

[13] R. Owczarzy, *Melting temperatures of nucleic acids: Discrepancies in analysis*, Biophys Chem **117** (2005) 207.

[14] V. Luzzati and A. Nicolaieff, *Etude par diffusion des rayons x aux petits angles des gels dacide desoxyribonucleique et de nucleoproteines - (note preliminaire).*, J Mol Biol **1** (1959) 127.

[15] C. Robinson, *Liquid-crystalline structures in polypeptide solutions.*, Tetrahedron **13** (1961) 219.

[16] Y. Bouligand, *Twisted fibrous arrangements in biological-materials and cholesteric mesophases.*, Tissue Cell **4** (1972) 189.

[17] F. Livolant, *Ordered phases of DNA in vivo and in vitro.*, Physica A **176** (1991) 117.

[18] M. H. Chow, K. T. H. Yan, M. J. Bennett, and J. T. Y. Wong, *Birefringence and DNA Condensation of Liquid Crystalline Chromosomes.*, Eukaryot Cell. **9** (2010) 1577.

[19] A. Minsky, E. Shimoni, and D. Frenkiel-Krispin, *Stress, order and survival.*, Nat Rev Mol Cell Biol **3** (2002) 50.

[20] F. Livolant and A. Leforestier, *Condensed phases of DNA: structures and phase transitions.*, Prog Polym Sci **21** (1996) 1115.

[21] T. Strzelecka, M. W. Davidson, and R. L. Rill, *Multiple liquid crystal phases of DNA at high concentrations.*, Nature **331** (1988) 457.

[22] F. Livolant, A. M. Levelut, J. Doucet, and J. P. Benoit, *The highly concentrated liquid-crystalline phase of DNA is columnar hexagonal.*, Nature **339** (1989) 724.

[23] G. Zanchetta and R. Cerbino, *Exploring soft matter with x-rays: from the discovery of the DNA structure to the challenges of free electron lasers.*, J Phys Condens Matter **22** (2010) 323102.

[24] H. H. Strey, J. Wang, R. Podgornik, A. Rupprecht, L. Yu, V. A. Parsegian, and E. B. Sirota, *Refusing to twist: demonstration of a line hexatic phase in DNA liquid crystals.*, Phys. Rev. Lett. **84** (2000) 3105.

[25] B. Tinland, A. Pluen, J. Sturm, and G. Weill, *Persistence length of single-stranded DNA.*, Macromolecules **30** (1997) 5763.

[26] Y. Lu, B. Weers, and N. C. Stellwagen, *DNA persistence length revisited.*, Biopolymers **61** (2002) 261.

[27] L. Onsager, *The Effects of Shape on the Interaction of Colloidal Particles.*, Ann. NY Acad. Sci. **51** (1949) 627.

[28] K. Merchant and R. L. Rill, *DNA length and concentration dependencies of anisotropic phase transitions of DNA solutions.*, Biophys. J. **73** (1997) 3154.

[29] P. Bolhuis and D. Frenkel, *Tracing the phase boundaries of hard spherocylinders.*, J Chem Phys **106** (1997) 666.

[30] J. V. Selinger and R. F. Bruinsma, *Hexagonal and nematic phases of chains. II. Phase transitions.*, Phys. Rev. A **43** (1991) 2922.

[31] A. M. Bohle, R. Holyst, and T. Vilgis, *Polydispersity and ordered phases in solutions of rodlike macromolecules.*, Phys Rev Lett **76** (1996) 1396.

[32] J. Piskur and A. Rupprecht, *Aggregated DNA in ethanol solution.*, Febs Lett **375** (1995) 174.

[33] J. J. Pelta, D. Durand, J. Doucet, and F. Livolant, *DNA mesophases induced by spermidine: structural properties and biological implications.*, Biophys J **71** (1996) 48.

[34] N. V. Hud and K. H. Downing, *Cryoelectron microscopy of lambda phage DNA condensates in vitreous Ice: The fine structure of DNA toroids.*, Proc Natl Acad Sci USA **98** (2001) 14925.

[35] T. Kuriabova, M. D. Betterton, and M. A. Glaser, *Linear aggregation and liquid-crystalline order: comparison of Monte Carlo simulation and analytic theory.*, J. Mater. Chem. **20** (2010) 10366.

[36] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, *Base-stacking and base-pairing contributions into thermal stability of the DNA double helix.*, Nucleic Acids Res **34** (2006) 564.

[37] C. Maffeo, B. Luan, and A. Aksimentiev, *End-to-end attraction of duplex DNA.*, Nucleic Acids Res. **40** (2012) 3812.

[38] M. R. Redinbo, L. Stewart, P. Kuhn, J. J. Champoux, and H. W. G., *Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA.*, Science **279** (1998) 1504.

[39] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, *Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution.*, J Mol Biol **319** (2002) 1097.

[40] L. Li, S. A. Pabit, J. S. Lamb, H. Y. Park, and L. Pollack, *Closing the lid on DNA end-to-end stacking interactions.*, Appl Phys Lett. **92** (2008) 223901.

[41] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, *Crystal structure analysis of a complete turn of B-DNA.*, Nature **287** (1980) 755.

[42] L. Campos, N. Valls, L. Urpí, and al, *Overview of the structure of all-AT oligonucleotides: organization in helices and packing interactions.*, Biophys. J. **91** (2006) 892.

[43] J. T. Davis and G. P. Spada, *Supramolecular architectures generated by self-assembly of guanosine derivatives.*, Chem Soc Rev **36** (2007) 296.

[44] G. Zanchetta, M. Nakata, M. Buscaglia, T. Bellini, and N. A. Clark, *Liquid crystal ordering of DNA and RNA oligomers with partially overlapping sequences.*, J. Phys. : Condens. Matter **20** (2008) 494214.

[45] T. Sarkar, C. C. Conwell, L. C. Harvey, C. T. Santai, and N. V. Hud, *Condensation of oligonucleotides assembled into nicked and gapped duplexes: potential structures for oligonucleotide delivery.*, Nucleic Acids Res **33** (2005) 143.

[46] C. De Michele, T. Bellini, and F. Sciortino, *Self-assembly of bi-functional patchy particles with anisotropic shape into polymers chains: theory, simulations and experiments.*, Macromolecules **45** (2012) 1090.

[47] C. De Michele, L. Rovigatti, T. Bellini, and F. Sciortino, *Self-assembly of short DNA duplexes: from a coarse-grained model to experiments through a theoretical link.*, Soft Matter **8** (2012) 8388.

[48] P. J. Collings, A. J. Dickinson, and E. C. Smith, *Molecular aggregation and chromonic liquid crystals.*, Liq Cryst **37** (2010) 701.

[49] P. Stano and P. L. Luisi, *Basic Questions About the Origins of Life: Proceedings of the Erice International School of Complexity (Fourth Course).*, Orig. Life Evol. Biosph. **37** (2007) 303.

[50] J. W. Szostak, D. P. Bartel, and P. L. Luisi, *Synthesizing life.*, Nature **409** (2001) 387.

[51] A. Pross, *How can a chemical system act purposefully? Bridging between life and non-life.*, J. Phys. Org. Chem. **21** (2008) 724.

[52] G. Schlesinger and M. S. L., *Prebiotic synthesis in atmospheres containing CH4, CO and CO2. I. Amino acids.*, J. Phys. Org. Chem. **19** (1983) 376.

[53] H. J. Cleaves, J. H. Chalmers, A. Lazcano, S. L. Miller, and J. L. Bada, *A Reassessment of prebiotic organic synthesis in neutral planetary atmospheres.*, Orig Life Evol Biosph **38** (2008) 105.

[54] S. J. Mojzsis, G. Arrhenius, K. D. McKeegan, T. M. Harrison, A. P. Nutman, and C. R. Friend, *Evidence for life on Earth before 3,800 million years ago.*, Nature **384** (1996) 55.

[55] E. Koonin, *How many genes can make a cell: the minimal-gene-set concept.*, Annu. Rev. Genomics Hum. Genet. **1** (2000) 99.

[56] D. P. Bartel and J. W. Szostak, *Isolation of New Ribozymes from a Large Pool of Random Sequences.*, Science **261** (1993) 1411.

[57] N. Paul and G. F. Joyce, *A self-replicating ligase ribozyme*, Proc. Natl. Acad. Sci. USA **99** (2002) 12733.

[58] H. S. Zaher and P. J. Unrau, *Selection of an improved RNA polymerase ribozyme with superior extension and fidelity*, RNA **13** (2007) 1017.

[59] Q. Vicens and T. R. Cech, *A natural ribozyme with 3',5' RNA ligase activity.*, Nat Chem Biol **5** (2009) 97.

[60] K. Ozawa, A. Nemoto, E. Imai, H. Honda, K. Hatori, and K. Matsuno, *Phosphorylation of nucleotide molecules in hydrothermal environments.*, Orig. Life Evol. Biosph. **34** (2004) 465.

[61] S. Miyakawa and J. P. Ferris, *Sequence- and regioselectivity in the montmorillonite-catalyzed synthesis of RNA.*, J. Am. Chem. Soc. **125** (2003) 8202.

[62] K. Grzeskowiak and L. Orgel, *Template-directed synthesis on short oligoribocytidylates.*, J. Mol. Evol. **23** (1986) 287.

[63] I. Budin and J. W. Szostak, *Expanding roles for diverse physical phenomena during the origin of life.*, Annu Rev Biophys. **39** (2010) 245.

[64] R. Pool and P. G. Bolhuis, *Prediction of an autocatalytic replication mechanism for micelle formation.*, Phys. Rev. Lett. **97** (2006) 018302.

[65] T. F. Zhu and J. W. Szostak, *Coupled growth and division of model protocell membranes.*, J. Am. Chem. Soc. **131** (2009) 5705.

[66] P. Baaske, F. M. Weinert, S. Duhr, K. H. Lemke, M. J. Russell, and D. Braun, *Prediction of an autocatalytic replication mechanism for micelle formation.*, Proc. Natl. Acad. Sci. USA **104** (2007) 9346.

[67] C. B. Mast, S. Schink, U. Gerland, and D. Braun, *Escalation of polymerization in a thermal gradient.*, Proc. Natl. Acad. Sci. USA **110** (2013) 8030.

[68] B. G. Bag and G. von Kiedrowski, *Templates, autocatalysis and molecular replication.*, Pure App. Chern. **68** (1996) 2145.

[69] G. Ertem, *Montmorillonite, oligonucleotides, RNA and origin of life.*, Orig. Life Evol. Biosph. **34** (2004) 549.

[70] J. P. Ferris and G. Ertem, *Oligomerization Reactions of Ribonucleotides on Montmorillonite: Reaction of the 5'-Phosphorimidazolide of Adenosine.*, Science **257** (1992) 1387.

[71] W. Huang and J. P. Ferris, *One-Step, Regioselective Synthesis of up to 50-mers of RNA Oligomers by Montmorillonite Catalysis.*, J. Am. Chem. Soc. **128** (2006) 8914.

[72] P. C. Joshi, S. Pitschb, and J. P. Ferris, *Homochiral selection in the montmorillonite-catalyzed and uncatalyzed Prebiotic synthesis of RNA.*, Chem. Commun. (2000) .

[73] A. Goldar and J. L. Sikorav, *DNA renaturation at the water-phenol interface.*, Eur. Phys. J. E **14** (2004) 211.

[74] A. I. Oparin, *The origin of life and the origin of enzymes.*, Adv. Enzymol. **27** (1965) 347.

[75] L. Onsager, *Life in the early days.*, Quantum Statistical Mechanics in the Natural Sciences (1974) .

[76] S. Rajamani, A. Vlassov, S. Benner, A. Coombs, F. Olasagasti, and D. Deamer, *Lipid-assisted Synthesis of RNA-like Polymers from Mononucleotides.*, Orig. Life Evol. Biosph. **38** (2008) 57.

[77] G. Costanzo, S. Pino, F. Ciciriello, and E. Di Mauro, *Generation of long RNA chains in water.*, J. Biol. Chem. **284** (2009) 33206.

[78] G. Zanchetta. PhD thesis, Milan, 2007.

[79] I. Dierking, *Textures of Liquid Crystals.* Wiley-VCH Verlag, 2003.

[80] G. Zanchetta, F. Giavazzi, M. Nakata, M. Buscaglia, R. Cerbino, N. A. Clark, and T. Bellini, *Right-handed double-helix ultrashort DNA yields chiral nematic phases with both right- and left-handed director twist.*, Proc. Natl. Acad. Sci. U.S.A. **107** (2010) 17497.

[81] R. Hentschke and J. Herzfeld, *Isotropic, nematic, and columnar ordering in systems of persistent flexible hard rods.*, Phys. Rev. A **44** (1991) 1148.

[82] S. Howorka, L. Movileanu, O. Braha, and H. Bayley, *Kinetics of duplex formation for individual DNA strands within a single protein nanopore.*, Proc. Natl. Acad. Sci. U.S.A. **98** (2001) 12996.

[83] W. J. Kim, T. Akaike, and A. Maruyama, *DNA Strand Exchange Stimulated by Spontaneous Complex Formation with Cationic Comb-Type Copolymer.*, J. Am. Chem. Soc. **124** (2002) 12676.

[84] H. Scher and E. W. Montroll, *Anomalous transit-time dispersion in amorphous solids.*, Phys. Rev. B **12** (1975) 2455.

[85] C. Monthus, *Anomalous diffusion, localization, aging, and subaging effects in trap models at very low temperature.*, Phys. Rev. E. **68** (2003) 036114.

[86] F. Topuz and O. Okay, *Rheological Behavior of Responsive DNA Hydrogels.*, Macromolecules **41** (2008) 8847.

[87] M. Fasolo, P. Sollich, and A. Speranza, *Phase equilibria in polydisperse colloidal systems.*, React. Funct. Polym. **58** (2004) 187.

[88] A. Speranza and P. Sollich, *Simplified Onsager theory for isotropic–nematic phase equilibria of length polydisperse hard rods.*, J. Chem. Phys. **117** (2002) 5421.

[89] E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli, and F. Sciortino, *Phase Diagram of Patchy Colloids: Towards Empty Liquids.*, Phys. Rev. Lett. **97** (2006) 168301.

[90] I. Budin and J. W. Szostak, *Expanding Roles for Diverse Physical Phenomena during the Origin of Life.*, Annu. Rev. Biophys. **39** (2010) 245.

[91] G. T. Hermanson, *Bioconjugate Techniques.* Academic Press, Elsevier, 2008.

[92] Z. A. Shabarova, N. G. Dolinnaya, V. L. Drutsa, N. P. Melnikova, and A. A. Purmal, *DNA like duplexes with repetitions. III. Efficient template-guided chemical polymerization of d(TGGCCAAGCTp).*, Nucl. Acids Res. **9** (1981) 5747.

[93] N. I. Sokolova, D. T. Ashirbekova, N. G. Dolinnaya, and Z. A. Shabarova, *DNA like duplexes with repetitions. III. Efficient template-guided chemical polymerization of d(TGGCCAAGCTp).*, FEBS Lett. **232** (1988) 153.

[94] N. G. Dolinnaya, N. I. Sokolova, D. T. Ashirbekova, and Z. A. Shabarova, *The use of BrCN for assembling modified DNA duplexes and DNA-RNA hybrids; comparison with water-soluble carbodiimide.*, Nucleic Acids Res. **19** (1991) 3067.

[95] W. S. Zielinski and L. E. Orgel, *Oligoanminonucleoside phosphoramidates. Oligomerization of dimers of 3'-amino-3'-deoxynucleotides (GC and CG) in aqueous solution.*, Nucleic Acids Res. **15** (1987) 1699.

[96] O. Taran. PhD thesis, Bochum, 2011.

[97] G. Von Kiedrowski, *A Self-Replicating Hexadeoxynucleotide.*, Angew. Chem. Inf. Ed. Engl. **25** (1986) 932.

[98] W. S. Zielinski and L. E. Orgel, *Autocatalytic synthesis of a tetranucleotide analogue.*, Nature **327** (1987) 346.

[99] M. Rossi, G. Zanchetta, S. Klussmann, N. A. Clark, and T. Bellini, *Propagation of chirality in mixtures of natural and enantiomeric DNA oligomers.*, Phys. Rev. Let. **110** (2013) 107801.

[100] N. P. Money, *Osmotic Pressure of Aqueous Polyethylene Glycols. Relationship between Molecular Weight and Vapor Pressure Deficit.*, Plant Physiol. **91** (1989) 766.

[101] H. H. Strey, V. A. Parsegian, and R. Podgornik, *Equation of State for DNA Liquid Crystals: Fluctuation Enhanced Electrostatic Double Layer Repulsion.*, Phys. Rev. Lett. **78** (1997) 896.

[102] P. J. Flory, *Molecular Size Distribution in Linear Condensation Polymers.*, J. Am. Chem. Soc. **58** (1936) 1877.

List of Publications

Refereed publications

T. Bellini, G. Zanchetta, T. P. Fraccia, R. Cerbino, E. Tsai, G. P. Smith, M. J. Moran, D. M. Walba and N. A. Clark, *Liquid crystal self-assembly of random-sequence DNA oligomers*, Proc. Natl. Acad. Sci. USA **109** (2012) 1110