A Liquid Crystal World at the Origin of Life

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Molecular and Cellular Biology

by MARCO TODISCO

Supervisors

PROFESSOR MARCO MUZI FALCONI

_____

PROFESSOR TOMMASO BELLINI

_____

February, 2020

SOMMARIO DELLA TESI

Il mistero dell'origine della Vita ha da sempre affascinato l'umanitá data la sua stretta connessione con un quesito fondamentale: *"da dove veniamo?"*

Per rispondere a questa domanda per centinaia di anni gli studiosi del campo hanno analizzato le condizioni presenti nel nostro Universo e sul nostro pianeta all'etá in cui la Vita é emersa, e testato se la Vita potesse nascere in uno scenario cosí complesso. Un percorso generale é stato definito per ricapitolare l'origine della Vita partendo da semplici molecole inorganiche fino a sistemi complessi, passando inevitabilmente attraverso il cosídetto *"mondo ad RNA"*; uno stadio nell'evoluzione della Terra in cui la Vita consisteva in lunghi polimeri lineari di RNA che traportavano informazione genica e catalizzavano la loro stessa replicazione, supportando un metabolismo primitivo.

Nonostante gli sforzi della comunitá scientifica, ad oggi non disponiamo ancora di una spiegazione su come lunghi biopolimeri si possano essere formati sulla Terra primordiale, data l'instabilitá dei legami fosfodiestere del RNA e visto che i modelli piú diffusi di polimerizzazione non enzimatica generano prodotti molto brevi. La necessitá di definire un percorso completo che abbia portato all'origine della Vita ci ha spinti a sviluppare un nuovo modello per guidare la formazione di un ribozima, basato sull'auto-assemblaggio di molecole di acido nucleico.

Sfruttando un ampio set di tecniche e concetti appartenenti ai campi della biologia molecolare, chimica e fisica della soft matter, in questo lavoro mostriamo come l'auto-assemblaggio e l'ordine sovramolecolare di acidi nucleici siano fenomeni con radici profonde nelle loro proprietá

chimico-fisiche, portando all'emergere di ordine liquido-cristallino in miscele di nucleosidi aventi diverse modificazioni chimiche e stato di polimerizzazione. La formazione di lunghi polimeri fisici con discontinuitá o mancanti completamente di un backbone, pur rimanendo strutturati seguendo la selettivitá di Watson e Crick, ci porta ad interrogarci su quale sia l'origine della celebre doppia elica, che sembra ora nascere dall'interazione di singoli nucleosidi.

In questa tesi verranno caratterizzati e descritti l'effetto dell'organizzazione sovramolecolare e della rottura di simmetria su brevi oligomeri di RNA e nucleosidi chimicamente attivati, mostrando come la loro reattivitá venga alterata in favore della formazione di lunghi polimeri lineari, sottolineando un nuovo potenziale percorso per l'origine della Vita.

THESIS ABSTRACT

The conundrum of the Origin of Life has always fascinated humankind, given its intimate relation with a fundamental question such as *"where do we come from?"*.

In order to answer this question, from centuries the scholars of the field have studied the conditions present in our Universe and on our planet at the age when Life emerged, and tested whether Life could raise in such a complex scenario. A general pathway has been defined to recapitulate the origin of Life from simple inorganic chemicals up to complex systems, and it inevitably passes through the so-called *"RNA World"*; a stage on our Earth when Life consisted of long linear polymers of RNA both carrying genetic information and catalyzing their very own replication and supporting a primitive metabolism.

Still, up to nowadays we lack a good explanation of how long biopolymers could have formed on the ancient Earth, given that RNA phosphodiester bonds are intrinsically unstable and that the most common models of non-enzymatic polymerization result in very short products. The necessity to define a complete pathway for the origin of Life has pushed us to develop a new model to drive the formation of a ribozyme, based on the self-assembly of nucleic acids molecules.

Exploiting a wide set of tools belonging to the fields of molecular biology, chemistry and soft matter physics, in this work we show that the self-assembly and supramolecular ordering of nucleic acids is a phenomenon with deep roots in their very physicochemical properties, leading to the emergence of liquid crystal ordering in mixtures of nucleosides having dif-

ferent chemical modifications and polymerization state. The formation of long physical polymers with discontinuities or completely lacking a backbone, while still structured following Watson-Crick selectivity, raises questions on the origin of the famous double helix, which seems now to be found in the features of single nucleosides interactions.

In this thesis the effect of supramolecular organization and symmetry breaking on short RNA oligomers and chemically activated nucleosides is described and characterized, showing that the chemical reactivity is altered favoring the formation of long linear polymers, and highlighting a new potential pathway for the Origin of Life.

Keywords: *Origin of Life*, *RNA World*, *Liquid Crystals*, *Supramolecular Chemistry*, *Ribozyme*.

ABSTRACT #1

Nonenzymatic Polymerization into Long Linear RNA Templated by Liquid Crystal Self-Assembly

Marco Todisco, Tommaso P. Fraccia, Greg P. Smith, Andrea Corno, Lucas Bethge, Sven Klussmann, Elvezia M. Paraboschi, Rosanna Asselta, Diego Colombo, Giuliano Zanchetta, Noel A. Clark, Tommaso Bellini

Self-synthesizing materials, in which supramolecular structuring enhances the formation of new molecules that participate to the process, represent an intriguing notion to account for the first appearance of biomolecules in an abiotic Earth. We present here a study of the abiotic formation of inter-chain phosphodiester bonds in solutions of short RNA oligomers in various states of supramolecular arrangement and their reaction kinetics.

We found a spectrum of conditions in which RNA oligomers self-assemble and phase separate into highly concentrated ordered fluid liquid crystal (LC) microdomains. We show that such supramolecular state provides a template guiding their ligation into hundred-bases long chains. The quantitative analysis presented here demonstrates that nucleic acid LC boosts the rate of end-to-end ligation and suppresses the formation of the otherwise dominant cyclic oligomers. These results strengthen the concept of supramolecular ordering as an efficient pathway toward the emergence of the RNA World in the primordial Earth.

This work has been published on ACS Nano[1] in 2018, was chosen by the editor for an invited commentary from Antonio Lazcano[2] and was awarded as one of the top ten papers published by a young member *(under 35)* of the Societá Chimica Italiana being a finalist for the Primo

Levi Prize 2018.

ABSTRACT #2

Backbone-free duplex-stacked monomer nucleic acids exhibiting Watson-Crick selectivity

Gregory P. Smith, Tommaso P. Fraccia, Marco Todisco, Giuliano Zanchetta, Chenhui Zhu, Emily Hayden, Tommaso Bellini and Noel A. Clark

We demonstrate that nucleic acid (NA) mononucleotide triphosphates (dNTPs and rNTPs), at sufficiently high concentration andlow temperature in aqueous solution, can exhibit a phasetransition in which chromonic columnar liquid crystal orderingspontaneously appears.

Remarkably, this polymer-free state exhibits, in a self-assembly of NA monomers, the key structural elements of biological nucleic acids, including: long-ranged duplexstacking of base pairs, complementarity-dependent partitioning ofmolecules, and WatsonCrick selectivity, such that, among all so-lutions of adenosine, cytosine, guanine, and thymine NTPs andtheir binary mixtures, duplex columnar ordering is most stable inthe A-T and C-G combinations.

This work has been published on PNAS[3] in 2018, was chosen by the editor for an invited commentary from Rudolf Podgornik[4] and was awarded the Cozzarelli Prize for Engineering and Applied Sciences.

Keywords: *Nucleic Acids*, *Prebiotic Evolution*, *Nucleoside Triphosphate*, *Double Helix*.

ABSTRACT #3

Phase separations, liquid crystal ordering and molecular partitioning in mixtures of PEG and DNA oligomers

Simone Di Leo, Marco Todisco, Tommaso P. Fraccia and Tommaso Belini

Liquid crystals (LCs) ordering of DNA and RNA oligomers relies on the presence of inter-duplex end-to-end attraction, driving the formation of linear aggregates. Such interactions are gauged, at a macroscopic level, by the osmotic pressure at the isotropic-nematic and nematic-columnar phase transitions.

We studied aqueous solutions of PEG and DNA duplex-forming oligomers, finding that there is a wide range of concentrations in which these mixtures phase separate into coexisting PEG-rich and DNA-rich phases, the latter being either in the isotropic state or ordered as a nematic or columnar LC. We determined the phase diagram in mixtures of PEG and DNA duplexes with different terminal motifs blunt ends, sticky overhangs, aggregation-preventing overhangs and measured the partitioning of the species in the coexisting phases.

On this basis, we determined the osmotic pressure as a function of the DNA concentration across the phase diagram. We compared the equation of state obtained in this way with both the CarnahanStarling equation of state for hard spheres and with the pressure predicted by computer simulations of a system of aggregating cylinders. We obtain a good agreement between experiments and simulations, and end-to-end attraction energies of the order of 6 kcal/mol, a bit larger than expected, but still in agreement with the current models for DNA-DNA interactions.

This work has been published on Liquid Crystals[5] in 2018.

Keywords: *equation of state*, *PEG-DNA phase separation*, *aqueous two-phase systems*, *DNA osmotic pressure*, *DNA self-assembly*, *Liquid Crystals*.

ABSTRACT #4

Liquid Crystal ordering of DNA Dickerson Dodecamer duplexes with different 5- Phosphate terminations

Marco Todisco, Gregory P. Smith and Tommaso P. Fraccia

The onset of liquid crystal (LC) phases in concentrated aqueous solutions of DNA oligomers crucially depends on the end-to-end interaction between the DNA duplexes, which can be provided by the aromatic stacking of the terminal base-pairs or by the pairing of complementary dangling-ends. Here we investigated the LC behavior of three blunt-end 12-base-long DNA duplexes synthesized with hydroxyl, phosphate and triphosphate 5-termini. We experimentally characterized the concentration-temperature phase diagrams and we quantitatively estimated the end-to-end stacking free energy, by comparing the empirical data with the predictions of coarse-grained linear aggregation models.

The preservation of LC ordering, even in presence of the bulky and highly charged triphosphate group, indicates that attractive stacking interactions are still present and capable of induce linear aggregation of the DNA duplexes. This finding strengthens the potential role of chromonic like self-assembly for the prebiotic formation of linear polymeric nucleic acids.

AIM OF THE THESIS

The content of this work is devoted to investigate and rigorously characterize the properties of nucleic acid liquid crystals, in order to provide evidence that the spontaneous supramolecular ordering of relatively simple biological molecules could have had a role in the origin of Life on the prebiotic Earth.

A broad set of problems have been tackled and are thoroughly described in the results session subdivided as follows:

1. Enhancement of chemical reactivity in short 5'-phosphate RNA oligomers is demonstrated through the analysis of ligation yield in mixtures with EDC, PEG and magnesium chloride. In this work a model for EDC-mediated ligation is developed to gain a better insight in the role of liquid crystal state in the enhancement of chemical reactivity. To accomplish this objective we have determined the multi-step chemical rates by modeling and numerically solving a set of differential equations, we have characterized the PEG-induced phase separation and measured the viscosity of our samples, obtaining a complete description of the system;

2. Suppression of circular products and intramolecular reactions is an emergent property of nucleic acids liquid crystals and their symmetry breaking. This property has been demonstrated by 2D PAGE and enzymatic digestion of ligation reactions products, in order to highlight molecules with different topology and a different electrophoretic mobility or resistance to enzyme degradation;

3. The addiction of terminal phosphate groups to short DNA oligomers destabilizes their liquid crystalline phases, while not completely disrupting them. The analysis and comparison of the phase diagrams of 5'-hydoxyl, 5'-monophosphate and 5'-triphosphate DNA oligomers through polarized optical microscopy reveals that supramolecular organization in nucleic acid is resilient to biologically relevant chemical modifications;

4. Liquid crystalline phases are found in mixtures of triphosphorylated single nucleosides. The study of ordered mesophases through polarized optical microscopy together with x-ray scattering data collected by our collaborators reveals that Watson-Crick pairing is at the basis of the formation of supramolecular structures in single nucleosides, removing the lower size limit for the formation of liquid crystals in nucleic acids and raising questions on the origin of the genetic code;

5. Ordered mesophases are found in concentrated mixtures of the chemically reactive 2-methylimidazole-activated guanosine. Its reaction products are characterized through phosphorus NMR and ESI mass spectrometry. A simple model for polymerization is shown here to account for our findings, pointing to a slow dynamic in the liquid crystal phases of G-quadruplexes.

6. The diffusivity of short oligonucleotides approaching the liquid crystal transition is measured through FRAP. These data set a new experimental approach to probe the dynamic of a system of small DNA fragments in a crowded environment.

The implications of these finding in the context of the origin of Life are

discussed and future perspectives and directions for this research topic are presented.

ACKNOWLEDGEMENT

ever required, together with an endless list of people that have been by my side during these years.

Lastly but most importantly, I would like to thank my biggest fans: my whole family, especially my mother **Maria Grazia** and my sister **Paola** for their constant moral support and guidance, along with patience and understanding.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TERMS AND ABBREVIATIONS

12BE   Self-complementary 5'-phosphate RNA molecule of sequence CGCGAAUUCGCG.

2MeImpG   2-methylimidazole-activated guanosine-5-monophosphate.

3pDD   Self-complementary DNA molecule of sequence CGCGAATTCGCG, with 5-triphosphate terminus and capability to produce liquid cyrstals.

6SE   Partially self-complementary 5'-phosphate RNA molecule of sequence GAUCGC.

ATP   Deoxyadenosine triphosphate.

C2   High order columnar liquid crystalline phase. COL Columnar liquid crystalline phase.

CTP   Deoxycytidine triphosphate.

DD   Self-complementary DNA molecule of sequence CGCGAATTCGCG, with 5-hydoxyl terminus and capability to produce liquid cyrstals.

EDC   *1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide*, a crosslinking chemical used in this work to produce novel phosphodiester bonds.

EDU   *1-Ethyl-3-(3dimethylaminopropyl)urea*, the byproduct of EDC reactions.

FRAP   Fluorescence Recovery After Photobleaching. A Microscopy technique used to probe a fluorophore diffusion.

FT   Partially self-complementary molecule of DNA of sequence

	TTCGCATGCGTT, with 5-hydoxyl terminus and TT overhangs disrupting end-to-end interactions and preventing formation of liquid crystals..
GMP	Deoxyguanosine monophosphate.
GTP	Deoxyguanosine triphosphate.
ISO	Isotropic state of matter.
LC	Liquid Crystal.
N*	Chiral nematic (cholesteric) liquid crystalline phase.
NA	Nucleic Acid.
nt	Nucleotides. Used in this work as a unit of measure for nucleic acid polymers length expressed as *number of monomers linked together*.
NTP	triphosphorylated nucleoside.
PAGE	Polyacrylamide gel electrophoresis. A technique used to separate and quantify nucleic acid molecules of different size and topology.
pDD	Self-complementary DNA molecule of sequence CGCGAATTCGCG, with 5-phosphate terminus and capability to produce liquid cyrstals.
pDNA	5'-phosphate DNA molecule.
PEG	Polyethylene glycol. Flexible linear polyether used in this work to induce phase-separation and liquid crystallization in nucleic acid mixtures.
pRNA	5'-phosphate RNA molecule.
rGMP	Guanosine monophosphate.

rTMP   Thymidine monophosphate.

sDD   Partially self-complementary DNA molecule of sequence GCGAATTCGCGC, with 5-hydoxyl terminus and capability to produce liquid cyrstals.

SVP   Snake Venom Phosphodiesterase I from *Crotalus adamanteus*, an enzyme with 3-exonucleolytic and endonucleolytic activities.

TT   Partially self-complementary molecule of DNA of sequence CGCGAATTCGCGTT, with 5-hydoxyl terminus and 3'-TT overhangs reducing end-to-end interactions and preventing the formation of liquid crystals.

TTP   Deoxythymidine triphosphate.

CHAPTER 1

Introduction

1.1 The Anatomy of DNA and RNA

1.1.1 The Chemistry of DNA and RNA

Nucleic acids are polymers of nucleotides, building units consisting of a pentose sugar (deoxyribose for DNA, ribose for RNA), esterified on the 5' position with a phosphate group and glycosylated in 1' position with a nitrogenous base among adenine (A), guanine (G), citosine (C), thymine (T) and uracil (U). The phosphorus atom in the nucleotide is highly electrophilic and easily reacts with nucleophiles such as the hydroxyl group of water or alcohols. The reaction with the hydroxyl group of a nucleotide with the phosphate group of a second one produces a thermodinamically unstable phosphodiester bond ($\Delta G = 5.3 kcal \cdot mol^{-1}$) and links the monomer into their biological polymeric form.
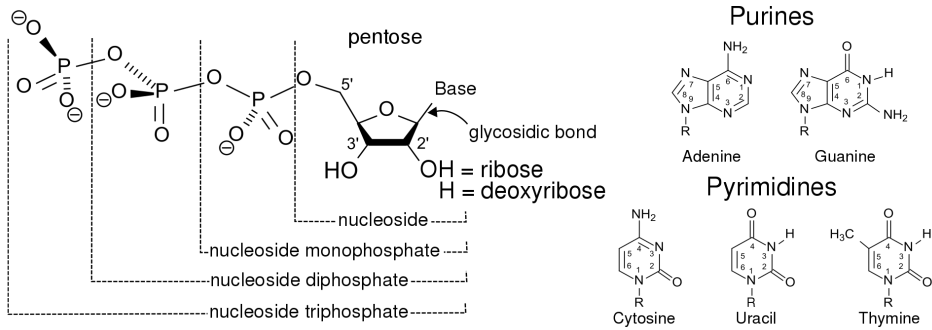
Fig. 1.1 Nucleotides structure.

Among DNA and RNA there are three main differences that must be taken into account for the purpose of this study:

i Ribose sugar adopts a different gometrical structure (C3'-endo) than deoxyribose (C2'-endo) to relieve the steric conflict given by eclipsed groups in a flat conformation and to reposition the phosphate group in a more distant position with respect to the oxygen in position 2'. This small effect propagates on large scales to a different geometrical structure of the macromolecule;[7]
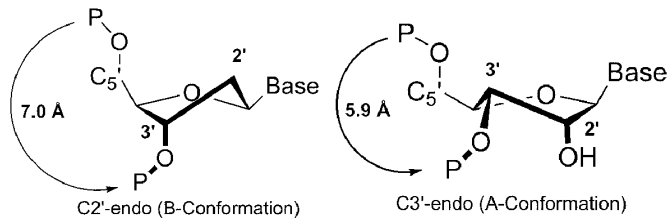
Fig. 1.2 DNA and RNA sugar pucker. The presence of a 2'-OH group in the ribose sugar modifies the overall conformation of the molecule and propagates up to its polymers.

ii Uracil is present in place of thymine. Uracil is a simpler and more soluble nitrogenous base and its substitution in DNA can be understood in view of evolution towards a safer genetic code, since uracil is indistinguishable from a deaminated citosine. When a deaminated citosine is found paired with a guanine, it is impossible for the cell

2

machinery to identifty which strand brings the wrong nitrogenous base and thus it cannot correct the issue without a 50% chance of introducing a genetic error;

iii Ribose sugar has a 2'-OH group which can attack the phosphate group in the 3' position in a RNA polymer after deprotonation, resulting in the cleavage of the strand leaving a 2',3'-cyclic phosphate group. This makes RNA intrinsically a much more unstable polymer with respect to DNA.

Nucleic acids behave as both acid and bases and undergo several ionizations. Since the nitrogenous bases are weakly soluble in water, the solubility of DNA is given by the backbone (sugar and phosphate) in their dissociated state. The phosphate backbone has such a low pKa that it can be always considered as completely ionized, making nucleic acids polymers *de facto* polyanions. Acidic DNA has a solubility that can easily be tenfold lower than basic DNA, as discussed more in detail in the *corpus* of the thesis. Nucleic Acids pKa are tabulated as follows, from Shabarova:[8]

<div align="center">**Table 1.1** Nucleic acid pKa</div>

Compounds	pKa Values		
	Nitrogenous Base	Pentose	Phosphate Group
Adenine	4.25		
Adenosine	3.63	12.35	
Deoxyadenosine	3.8		
Adenosine 5'-P	3.74	13.06	0.9; 6.05
Guanine	3.0; 9.32		
Guanosine	2.1; 9.33	12.3	
Deoxyguanosine	2.4; 9.33		
Guanosine 5'-P	2.9; 9.6		0.7; 6.3
Uracil	7.48; 11.48		
Uridine	9.25	12.59	
Deoxyuridine	9.3		
Uridine 5'-P	9.5		1.0; 6.4
Cytosine	4.6; 12.2		
Cytidine	4.1	12.24	
Deoxycytidine	4.25		
Cytidine 5'-P	4.5		0.8; 6.3
Thymine	9.94		
Deoxythymidine	9.8	12.85	
Deoxythymidine 5'-P	10.0		1.6; 6.5

1.1.2 The Physics of DNA and RNA

1.1.2.1 Structural Properties

Even though the antiparallel Watson-Crick base pairing is the bonding configuration found in living beings chromosomes, many possible configurations are possible[9] and are seldom found stabilizing tertiary structures or non-canonical motifs such as G-quadruplexes. The possible hydrogen bonding configurations can be summarized as depicted in Fig. 1.3.

Glycosidic Bond Orientations
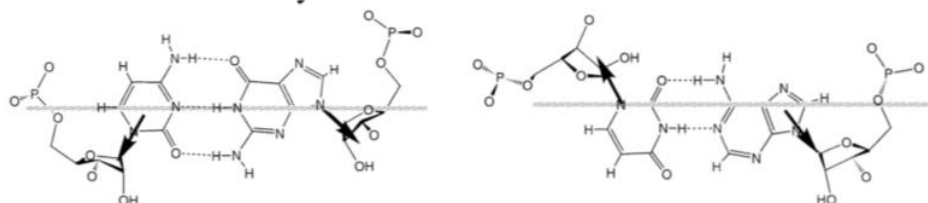
Cis orientation of the Glycosidic Bonds   Trans orientation of the Glycosidic Bonds

Interacting Edges
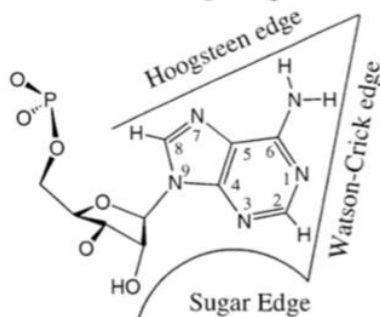
No.	GLYCOSIDIC BOND ORIENTATION	INTERACTING EDGES	DEFAULT LOCAL STRAND ORIENTATION
1	Cis	Watson-Crick / Watson-Crick	Anti-parallel
2	Trans	Watson-Crick / Watson-Crick	Parallel
3	Cis	Watson-Crick / Hoogsteen	Parallel
4	Trans	Watson-Crick / Hoogsteen	Anti-parallel
5	Cis	Watson-Crick / Sugar Edge	Anti-parallel
6	Trans	Watson-Crick / Sugar Edge	Parallel
7	Cis	Hoogsteen / Hoogsteen	Anti-parallel
8	Trans	Hoogsteen / Hoogsteen	Parallel
9	Cis	Hoogsteen / Sugar Edge	Parallel
10	Trans	Hoogsteen / Sugar Edge	Anti-parallel
11	Cis	Sugar Edge / Sugar Edge	Anti-parallel
12	Trans	Sugar Edge / Sugar Edge	Parallel

Fig. 1.3 Possible hydrogen bonding configurations, from Leontis et al.[9]

In the antiparallel Watson-Crick configuration, the nitrogenous bases A and T (or A and U) can produce two hydrogen bonds, while C and G produce three, making the interaction considerably stronger. It is important to notice that most of the energetic contribution for the formation of a double helix is given by stacking, while base pairing is either irrelevant (for G-C) or even destabilizing the interaction (for A-T),[10] but working as a gatekeeper for the specificity of the base pairing.

Since the nature of the interaction between two nucleic acid strands is well known, a model has been developed and perfected over the years to calculate the ΔG associated with the formation of a double stranded nucleic acid starting from two single strands, using just their sequences as an input for the prediction.[11] This model relies on quadruplets as elementary units for the calculation, meaning that the minimum element containing the information for the stability of the double stranded structure is given by a paired nucleotide and its neighbours. This feature of the model is necessary to take into account the contribution of stacking interaction, that would be otherwise neglected just by looking at the sequence composition from a base-pair level. The energetic contribution to the formation of double helices and more exotic motifs have been characterized with such a great detail that it is possible nowadays to design *a priori* a set of multiple sequences to self-assemble in the desired structure with a chosen stability and nanometric precision.[12]

The hydrogen bonding between two sufficiently long nucleic acid polymers results in the formation of a double helix. In principle one could expect the formation of a "ladder-like" polymer, which is never observed. In order to understand why this is the case, we must keep into account the hydrophobicity of the nitrogenous bases, that induces their collapse through hydrophobic stacking interaction in order to reduce the area exposed to the water solvent. The sugar-phosphate backbone has a fixed phosphate-to-phosphate distance of $\approx 6 \mathring{A}$, restricting the collapse of nitrogenous bases,

whose distance could be in principle reduced up to 3.4Å. To minimize their distance, a tilt is necessary, which can be propagated as an helix with a $\approx 2nm$ diameter, whose features can be easily recovered with a simple geometric construction.
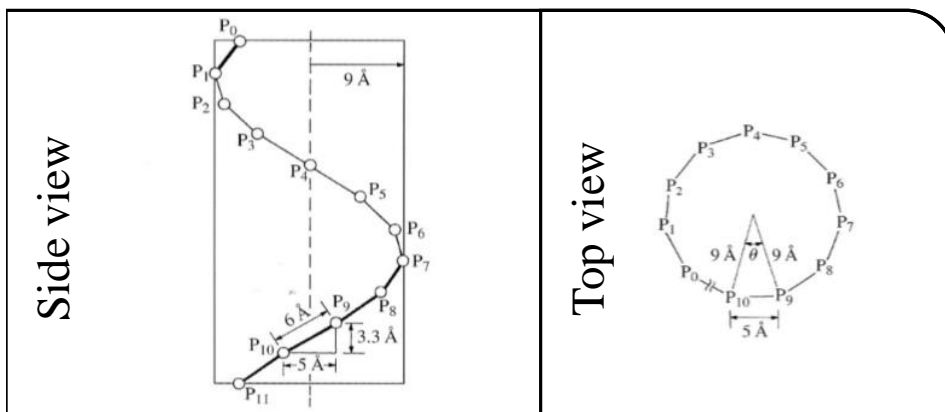
Fig. 1.4 NA double helices features can be recovered by a simple geometrical construction. (a) Phosphates are represented by circles and sugars by thin lines; (b) top view, from *Understanding DNA*.[13]

It is important to keep into account that the presence of a 2'-hydroxyl group in ribose alters the details of the double helix produced by the RNA, the so called A-form, with respect to the one produced by DNA, the so-called B-form (see Fig. 1.2), whose features are as tabulated:

Table 1.2 Nucleic acid helix properties

	A-DNA	B-DNA
Helix handedness	Right	Right
bp per turn	11	10.5
Distance between bp	2.9Å	3.4Å
Tilt with the respect to main axis	+19°	-1.2°
Diameter	25.5Å	23.7Å
Sugar conformation	C2'-endo	C3'-endo

The aromatic character of the nitrogenous bases makes them planar and confers them a strong UV absorbance, routinely used both to quantify the amount of nucleic acid in used samples and to investigate their structural properties.
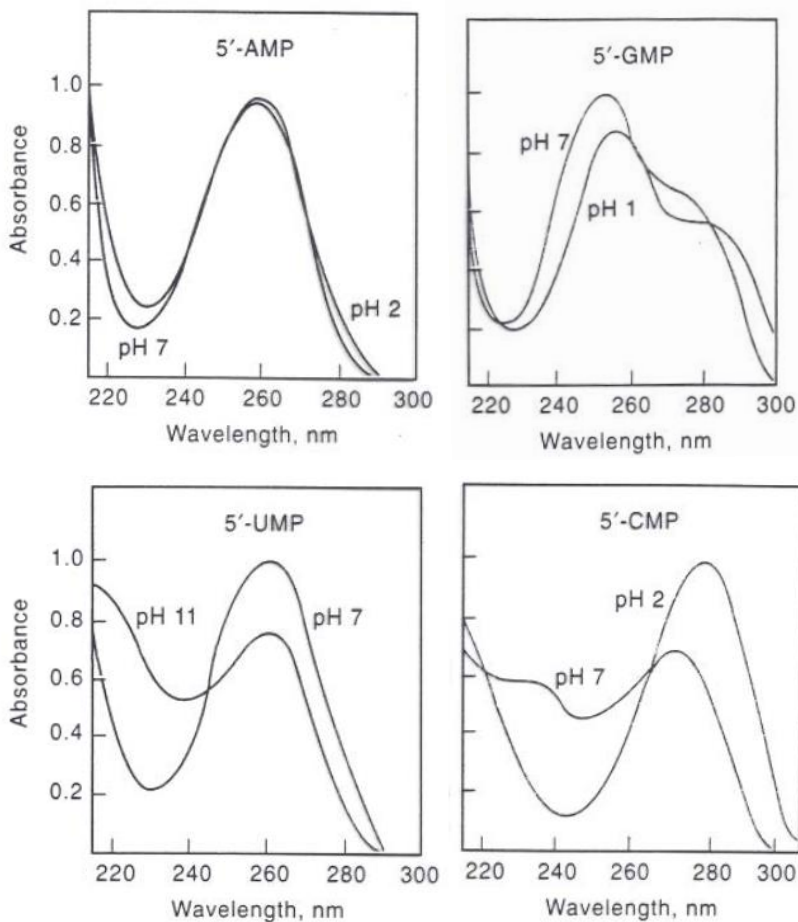
Fig. 1.5 Ribonucleotides UV absorbance spectra.

The pairing of the nucleic acid strands through hydrogen bonding, limiting the excitation delocalization along the nitrogenous bases,[14] and influencing their dipole-dipole interaction,[15] causes a reduction in absorbance, the so-called hypochromic effect of nucleic acids. The measurement of the UV absorbance of nucleic acid samples over temperature allows to

extrapolate structural and thermodynamical information[11] thanks to the hypochromic effect.

For the purposes of this work, a short double helix of DNA or RNA will be considered as a rigid rod-like structure. The persistence length of a polymer (P) is defined as the the characteristic length at which the orientations of the polymer at two points become uncorrelated. More precisely, taken an arbitrary point along the polymer and defined its orientation as $\theta = 0$, let's consider the orientation θ of the polymer in a different location at distance a L. The cosine of the angle θ averaged over all starting positions can be expressed as follows:

$$\langle cos\theta \rangle = e^{-\frac{L}{P}} \tag{1.1}$$

From an empirical point of view, a polymer tract of a length below P behaves as a rigid rod, while above a length P is flexible. For a single stranded nucleic acid the persistence length varies depending on the ionic strength[16] from $\approx 1nm$ to $2nm$, a range comparable with the distance between two adjacent nucleotides, but it raises to an astounding $\approx 50nm(150bp)$ for a double helix. While 150bp seem very few compared to the size of a genome, they are comparable to the length of exons and introns as well as the nucleosomic DNA wrapped around histons, and it is a size much larger than the one of the molecules that will be used in this work,.

1.1.2.2 Entropy driven phase separation

It has been shown that very short double helices of DNA or RNA, when mixed with high concentrations of polyethilene glycol (PEG), can phase separate in two compartments, a DNA-rich one and a PEG-rich one.[17,5] These two phases are fluid and molecules can diffuse between them, making them suitable for our experimental purposes as discussed in Chap. 2.1.

To understand how short nucleic acid double helices and PEG do phase separate we have to look into the theory behind these processes. Given a

binary mixture of non-interacting compounds A and B, the free energy of the system is reduced when a homogeneous solution is produced for every combination of the two components. In other words, the Gibbs free energy of mixing, defined as $\Delta_{mix}G = \Delta_{mix}H - T\Delta_{mix}S$ is always negative, with a global minimum found when the relative volume fraction of one component (x) is equal to 0.5, i.e. the system is balanced.

This might not be true for a system of interacting particles. In a oil and water mixture for example, the mixed system is not stable for $x = 0.5$, it will thus partition in two separate phases. The Gibbs free energy of mixing of such systems have a more complex shape with two minima at x' and x" and an unstable (or metastable) region in-between. When the composition of the system is in the unstable region, the mixture decomposes in two phases having compositions $x = x'$ and $x = x''$.

(a) Miscible System

$\Delta F_{\mathrm{mix}}/NkT$
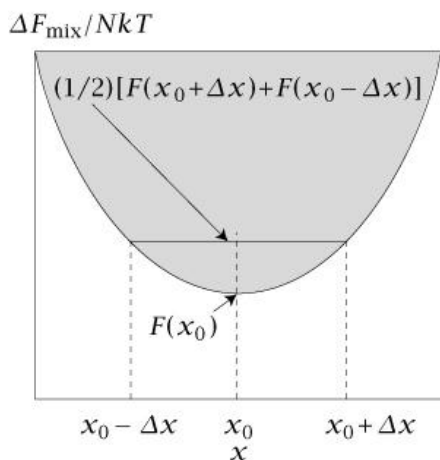
(b) Immiscible System

$\Delta F_{\mathrm{mix}}/NkT$
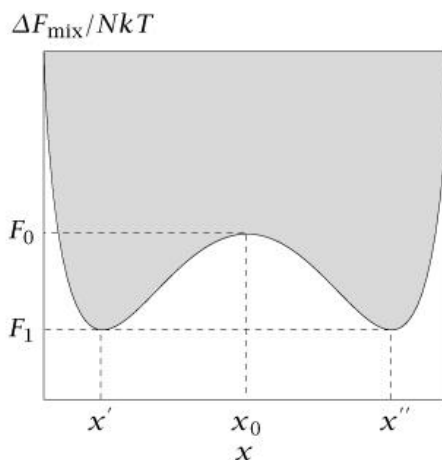
Fig. 1.6 Free energy of mixing as a function of volume fraction in miscible and immiscible systems.

For a mixture of flexible and rigid molecules such as PEG and RNA (or DNA) double helices, the system is driven to mixing by the self-repulsion between negatively charged nucleic acid molecules, but a larger free energy contribution due to entropy increase is produced when the system

demixes. The classic Asakura-Oosawa model[18] explains this phenomenon introducing depletion forces arising from excluded volume effects.

While flexible molecules such as PEG can interpenetrate, double helices of nucleic acids are rigid rods and thus cannot get closer to the center of mass of other nucleic acids double helices or PEG molecules than their radius of gyration. This produces an excluded volume around nucleic acids molecules which is a part of the space not available for diffusion of other species. Bringing in close proximity two nucleic acid molecules reduces their excluded volumes by their volume of overlap and increases the number of configurations that can be assumed by PEG molecules, producing an effective driving force towards the phase separation of the system.

This feature is particularly relevant if we consider that it is due to a simple difference in flexibility, and can be extended to other flexible polymers aside PEG, such as single stranded nucleic acid molecules or prebiotic crowding agents.

1.1.3 Chemical Reactivity of Nucleic Acids

The main chemical reactions taken into account in this study are phosphoryl transfer reactions, which can be described as a concerted S_n2:
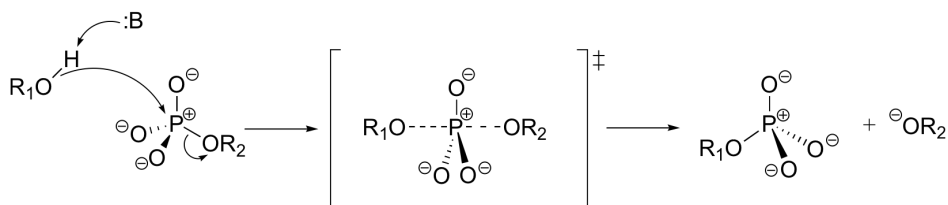
Fig. 1.7 Mechanism for S_n2 reaction involving phopsphoryl transfer.

It is important to notice that whether biological phosphoryl transfer reactions follow a concerted S_n2, an *addition-elimination* process or a S_n1 is still a matter of debate and dependent on substrates and the details of the enzymes active site.

For the scheme proposed above, as the nucleophile approaches the

phosphorus atom, it changes its bonding geometry from tetrahedral to trigonal bipyramidal, becoming what is defined as the transition state. In this transition state the phosphorus atom has to delocalize a great amount of electronegative charge, phenomenon described in the Arrhenius law in the *entropy of activation* term. This is usually accomplished in biological systems by the use of Lewis acids such as one or more magnesium cations in the active site of an enzyme, that coordinate the non-bridging oxygen atoms of the phosphate group catalyzing the reaction.[19]

Depending on the nucleophile group and the reaction order, a phosphoryl transfer reaction can describe a set of phenomena occurring in nucleic acids samples such as the followings:

Table 1.3 Nucleic acid reactivity

Reaction Order	Nucleophile	Product
First Order	RNA 2'-OH	2',3' cyclic NA
First Order	DNA or RNA 3'-OH	3',5' cyclic NA
Second Order	Water OH	Cleavage
Second Order	RNA 2'-OH	Phosphodiester Bond
Second Order	DNA or RNA 3'-OH	Phosphodiester Bond
Second Order	ImpddN 2'-NH_2	Phosphoramidate Bond

Over the years, attempts to increase the reaction yield for the formation of nucleic acids polymer has pushed the research towards more reactive and less charged leaving groups than polyphosphates, such as imidazole, methylimidazole and 2-aminoimidazole, still largely used and focus of modern scientific activity.

1.2 Nucleic Acids Liquid Crystals

All existing matter is organized in a state such as solid, liquid or gas, each featuring a certain degree of positional and orientational order. If we take a liquid, we cannot find long-range positional and orientational order, while in a crystalline solid, these are well-defined. In a liquid crystal (LC) the molecules can flow, as in a conventional liquid, while retaining some degrees of positional and orientational order.
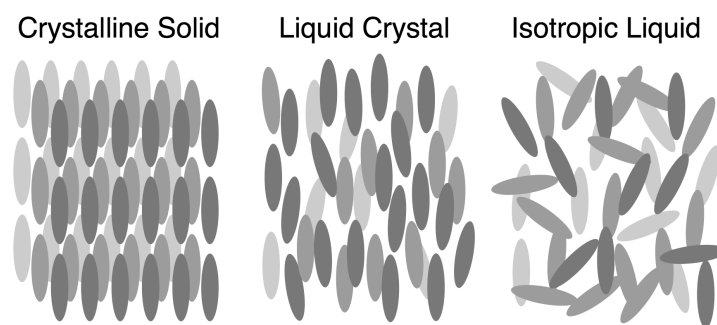
Fig. 1.8 Liquid crystal state of matter.

A peculiar property of liquid crystals is their birefringence; the capability of these molecules to slow down the interacting electromagnetic wave depending on its polarization state. This property is commonly exploited to define whether the solution we are working with is ordered or not: if an isotropic sample is observed between two crossed polarizers, no light gets transmitted. Instead, if the sample is birifringent light is transmitted and peculiar textures can be observed reflecting the underlying supramolecular order.

The formation of lyotropic LCs is usually dominated by the entropic contributions of the system, controlled by the highly anisotropic molecular shape. These give rise to supramolecular order in pure substances (thermotropic LCs) or in water solutions (lyotropic LCs). Nucleic Acids Liquid Crystals belong to this last class.

The tendency of DNA to produce liquid crystals has been known since a long time. For example, birifringent textures were observed in solutions of the highly monodispersed DNA purified from nucleosomes cores, 146bp in length.[20]

This transition to a DNA liquid crystalline state is controlled by concentration, and can be understood modeling the system as a solution of rigid rods with no interactions other than steric repulsion (meaning that the rods cannot interpenetrate). This system has been widely studied analytically by Lars Onsager,[21] who defined a threshold concentration for the formation of supramolecular order at a volume fraction occupied by the rigid rods $\phi \approx 4D/L$, where D is the rod diameter and L is its length.

According to this theory, no liquid crystal phase should be obtained for a DNA double helix length shorter than 28bp, nevertheless, it has been observed for much shorter oligomers thanks to the contribution of end-to-end interactions[22] leading to the formation of long linear aggregates that satisfy Onsager requirements for the transition to a liquid crystalline state.[23,24] Two different phases are usually observed in nucleic acid liquid crystals.

1.2.1 Nematic Liquid Crystals

In a nematic liquid crystals the linear aggregates or strongly anisotropic molecules acquire a common orientation in space along the director vector **n**. The centers of mass of the molecules are not correlated and a measurement of their alignment can be defined using the order parameter (S):

$$S = \frac{1}{2} \langle 3cos^2\theta - 1 \rangle \tag{1.2}$$

Where θ is the angle between the major molecular axis and the director vector. S can vary from 0 for completely unaligned (isotropic) liquid to 1 for a perfectly aligned nematic liquid crystal.

If the builiding units of the liquid crystals are chiral molecules as DNA and RNA helices, a macroscopic helical superstructure with pitch p can be

superimposed on the nematic phase, with a twist axis perpendicular to the director, which is now defined as local due to its continuous twist. This new phase is called chiral nematic of cholesteric (N*).
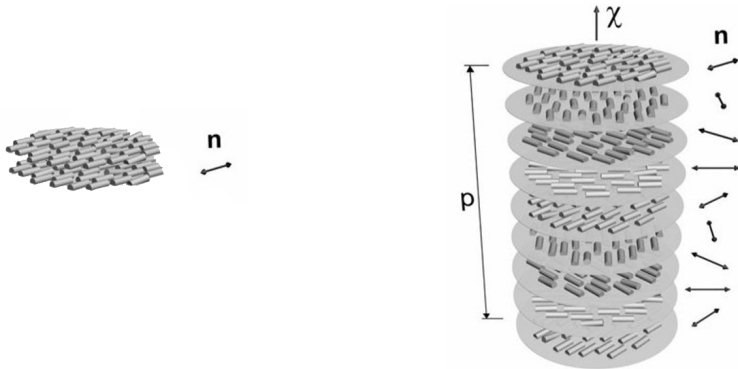
Fig. 1.9 On the left representation of a nematic LC. The molecules are on average aligned with the nematic director **n**. On the right representation of a chiral nematic LC. The molecules are locally aligned with a nematic director **n** that twists over a perpendicular axis χ. Figure adapted from *dept.kent.edu*

Given that a structure with a periodicity p is present in a cholesteric phase, Bragg reflection on visible light can occur. If a beam of coherent light shines over a grating of spacing d with an incoming angle θ, the light component of wavelength λ will be constructively reflected, where λ is defined as $2d \cdot sin\theta$.

In a cholesteric liquid crystal sample, the macroscopic helical super-structures produces a grating of spacing $p = 2d$, thus light of wavelength $\lambda = p$ will be reflected at an angle $\theta = 90°$, giving a direct measurement of submicron pitches.
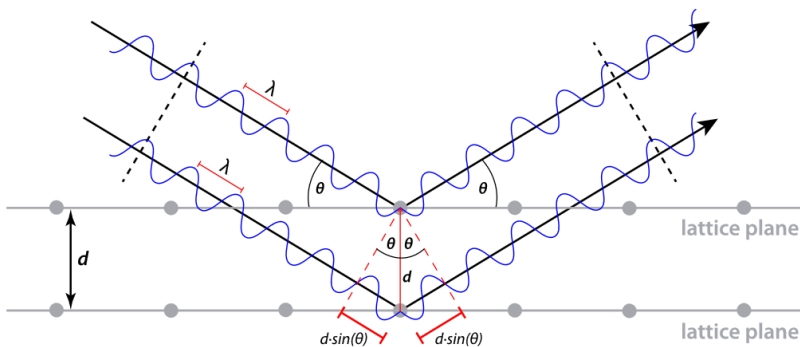
Fig. 1.10 Representation of constructive interference of a coherent wave of wavelength λ on a grating with spacing d. Figure adapted from *https://www.didaktik.physik.uni-muenchen.de*
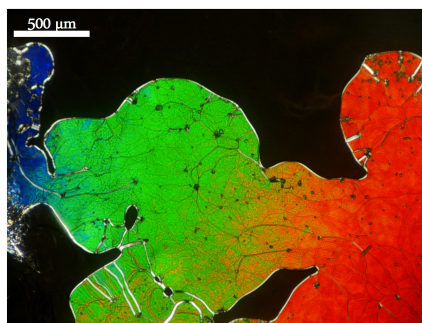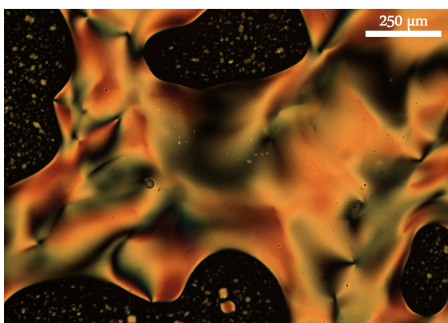
Fig. 1.11 Examples of a nematic (left) phase produced by 2MeImpG and a cholesteric (right) phase produced by a 12bp double helix of RNA. Notice the array of colors in the cholesteric phases, corresponding to submicron pitches in the macro helical superstructure.

1.2.2 Columnar Liquid Crystals

A columnar liquid crystal of nucleic acids consists in stacked aggregates that organize in long columns aligned on a hexagonal 2D lattice. A remarkable difference between a hexagonal packing crystal is that the columns of a columnar liquid crystal (COL) bend over long distances and can slide one to respect to the other. This features is reflected in the texture of the focal conics, structures where the optical axis of the liquid crystal changes con-

16

tinuously over hundred of micrometers creating dark areas when aligned with the polarizing filters (see Fig. 1.10a).
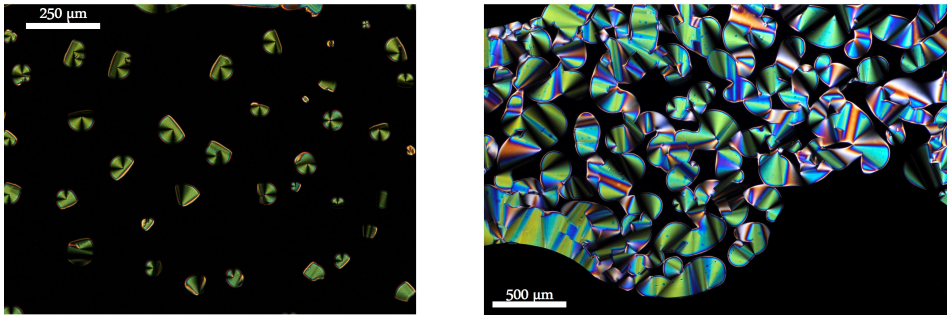
Fig. 1.12 Examples of columnar phases produced by a 12bp double helix of RNA (left) or DNA (right).

1.3 The Origin of Life

1.3.1 The Biogenesis

"Where do we come from?" is an intimate question that always accompanied human kind. Such a question evokes answers from the religious and metaphysical domains of knowledge, but Science has its share of things to say on the topic.

Going back to the XVIII century, it was still commonly accepted the Aristotelian notion that Life could emerge from inanimate material. This idea was raising from the observation of creatures such as fleas and maggots apparently rising from dust and dead flesh, a misconception given by the lack of a proper metholody and framework to evaluate the hypothesis.

Italian naturalists Francesco Redi and Lazzaro Spallanzani tested this hypothesis in the XVII and XVIII centuries with solid scientific approach and proper controls; the first showing that maggots could emerge only when a sample of rotting flesh was accessible and not when it was covered with a gauze. As a complementary experiment, Lazzaro Spallanzani could show that a high-temperature treatment inhibited the formations of maggots, showing that the production of life was not an inherent feature of matter. The final blow to this centuries-long misconception was given in the XIX century by Louis Pasteur, who designed an ingenious experiment to confute the objections posed to his predecessors, such as heat inhibiting the "vegetative force" of matter.
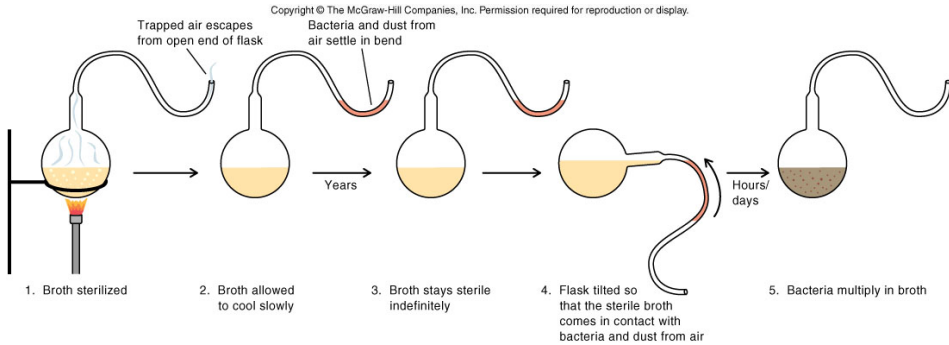
Fig. 1.13 Louis Pasteur experiment to confute abiogenesis.

The concept of *"Omne vivum ex vivo"* became stronger than ever before, while abiogenesis, Life emerging from inanimate matter, seemed to be a chimaera.

1.3.2 The Abiogenesis

In 1859 Charles Darwin published his book *On the Origin of Species*, exposing what we call today as Darwinian Evolution. The work clearly stated the role of natural selection in guiding the modification of every living being over million of years, tracing similarities and progressions between relative species across the globe.
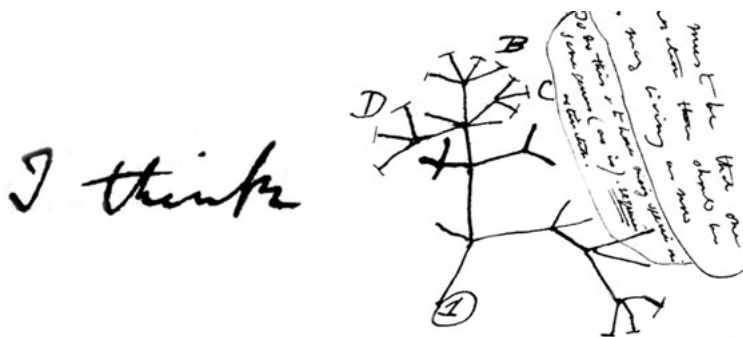
Fig. 1.14 Charles Darwin sketch made in 1837.

The natural consequence of the design of a branched tree with a common root was to point out the existance back in time of a shared ancestor

among all living beings, and indeed Charles Darwin grasped this concept.

Therefore I should infer from analogy that
probably all the organic beings which have ever
lived on this earth have descended from some one
primordial form, into which life was first breathed.

To trace back in time the aspect of the first living being, nowadays called Last Universal Common Ancestor (LUCA), it is necessary to simplify the features common and necessary among the current known living beings, a tough task that requires a deep knowledge of molecular biology and organic chemistry, both missing at the time.

Nevertheless, the scientific community put its effort to solve such an intricate conundrum, leading to the first solid scientific approach to a natural origin of Life taken a century ago by Aleksandr Oparin, whose work had been focused on extending back in time Darwininan Evolution up to lifeless matter. Oparin proposed coacervates as the self-organized state of the first protocelles assembled from the organic molecules present in the prebiotic soup. The development of Organic Chemistry together with Oparin's effort were of pivotal importance in opening a new field of research, celebrated by the very first meeting of the origin of Life organized in Moscow in 1957.[25]

Still, Oparin's theory could not stand the test of time and among the others it was bugged by a deep focus on coacervates instead of liposomes and the lack of knowledge of molecular biology and the central role of nucleic acids. Regardless of Oparin's contribution, today we consider our work as the legacy of an experiment performed in 1952 by a PhD Student enrolled at the University of Chicago in the laboratory of the Nobel Prize laureate Harold Urey.

1.3.3 The Miller-Urey Experiment

The idea behind the *Miller-Urey Experiment* was to provide energy, in the form of electric discharge, into a mixture reproducing the conditions of the ancient atmosphere of our planet as people believed those to be back in 1952.

Even though along the years there has been a lot of dispute regarding the exact composition and the oxidizing or reducing nature of the prebiotic atmosphere, the experiment showed to the scientific community that organic molecules such as aminoacids could have been produced under completely abiotic conditions.

Nowadays, many variants of the experiment have been performed and even the synthesis of ribonucleotides is no longer an obstacle[26] and a plethora of chemicals have been found in outer space and in carbonaceous meteorites.[27,28]

A natural explanation for the origin of Life seems to be graspable by human kind.

1.3.4 The RNA World hypothesis

Since the production of the building blocks of living beings started to be more and more clear, the problem shifted to the research of a path that could bring from the simple organic molecules produced in the *Miller-Urey Experiment* and similar experiments to a self-replicating entity capable to evolve into a complex living being. A game changer in the field has been the discovery of ribonucleic acid polymers capable of catalyzing chemical reactions.

These molecules, single linear strands that could fold assuming a specific and functional three-dimensional shape, where baptised ribozymes and their impact on the community was so relevant that their discovery valued a Nobel Prize in Chemistry to Thomas Cech.

The Central Dogma of Molecular Biology states that genetic information is stored in deoxyribonucleic acid, used to encode ribonucleic acid and then translated into a specific aminoacid sequence that composes a target protein. Proteins allow Life to be, catalyzing chemical reactions and the processing of the Central Dogma itself, solving the complex task of replicating, transcribing and translating the genetic code.

Scientists were put in front of an amazing problem: since nucleic acids and proteins exist nowadays in such an entwined way that we cannot think of a system in which we just get rid of one of them, if Life has originated with the same Central Dogma as today, than nucleic acids and proteins must have raised and coexisted at the same time.

To produce at the same time two complex polymers such as DNA and proteins seems an impossible task, what seems possible instead, is to produce an *RNA World* dominated by a single species of polymer - ribonucleic acid - both carrying the genetic information and catalyzing its own replication and metabolism. This concept, first proposed by Carl Whoese, Francis Crick and Leslie Orgel, has stood the test of time and is still relevant, even if some downsides affect its feasibility.

The instability of the phosphodiester bond in water and the intrinsic hydrolysis mechanism of RNA pose a great obstacle towards the formation of the RNA World. In order to produce a ribozyme one has to either stabilize the chemical bonds to reduce degradation or to kinetically overcome it increasing the reactivity of the nucleotides. Following this direction, James Ferris has shown that a solid support such as the Montmorillonite clays, mineral deposits produced by the weathering of volcanic ashes, can adsorb single nucleotides and catalyse their polymerization,[29] while other scientists such as Jack Szostak have shown that short single strands of nucleic acids can behave as templates to enhance the reactivity of transitorially paired nucleotides that would otherwise be very weakly reactive or completely unreactive.[30,31,32]

Still, the ribozymes evolved in laboratory to catalyse the transcription of a RNA substrate[33] or their self-replication[34] require a length in the order of the hundreds of nucleotides to properly fold and be functional. In contrast, the length of the polymers produced non-enzymatically is usually quite limited.

1.3.5 Alternative Hypotheses

Alternative scenarios have been proposed over the years with the aim of overcoming the issues connected with the *RNA World* hypothesis and the idea of a living system produced hierarchically after the first replicator was established.

In the context of the study of complex systems, the emergence of a self-replicating unit from a network of interacting chemicals was first proposed by Alexandr Oparin, who did not have any knowledge of modern Molecular Biology but could see how Life could emerge as a collective property of a molecular mixture. This pathway, that has been more deeply and extensively studied in modern times,[35] has been baptized as Metabolism-First hypothesis.

In a scenario in which a set of chemicals connected by chemical reactions catalyzed by a subset of the chemicals present in the system, an autocatalytic cycle can be established and, in principle, fullfill NASA definition of life as a "self-sustaining chemical system capable of Darwinian evolution", even if a lot of debate is still present whether these compositional genomes are indeed compatible with a pathway towards the formation of living beings.[36]
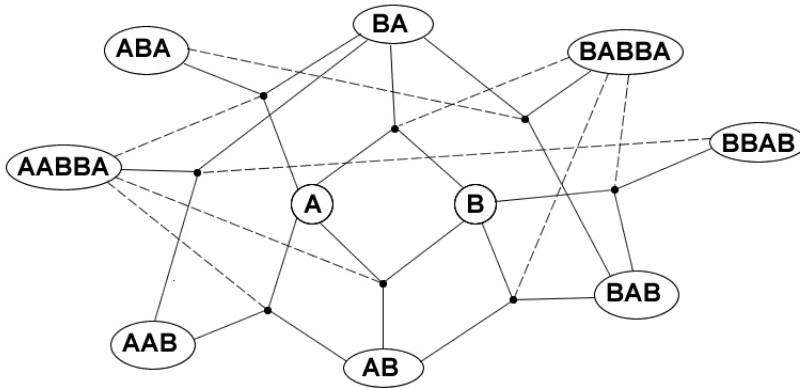
Fig. 1.15 Example of an autocatalytic system as described by Stuart Kauffman

Moreover, aside RNA replicators, scientists are exploring the possibility of a Amyloid-First[37] and Protein-First hypothesis as well,[38] exploiting the catalytic activity of supramolecular assemblies of short oligopeptides or patches exposed by foldamers.

1.4 A Liquid Crystalline Origin of Life

Given the observation of liquid crystalline phases in mixtures of short oligomers of DNA, a natural evolution of this research was to understand whether the physical proximity between their termini could influence their chemical reactivity.

According to De Michele et al.[23] the average aggregate size (M) in a coarse-grained model of quasi-cylinders with sticky ends exhibit a discontinuity when the system transitions from isotropic to nematic phase.
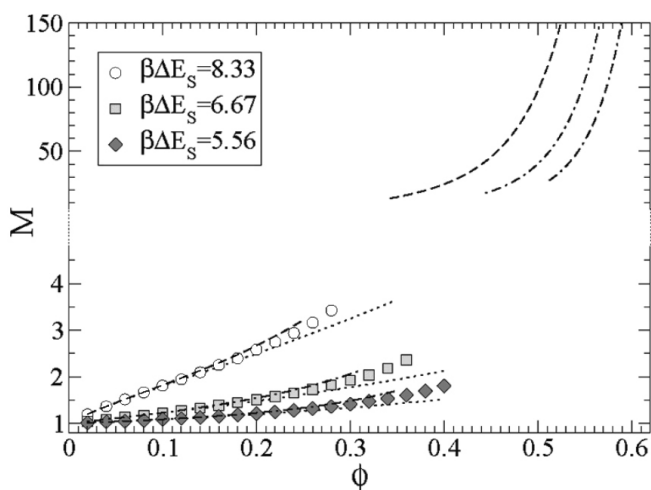
Fig. 1.16 Average aggregate size dependence on nucleic acid concentration, figure adapted from De Michele et al.[23]

The presence of such a discontinuity implies that the supramolecular order produced by the aggregates in return does stabilize the aggregates. In a liquid crystal solution we can thus expect the presence of very long *physical polymers*, which can be defined as linear aggregates of a starting double helix with no covalent bonds holding them together.

Following the idea that weak bonds at the basis of supramolecular assemblies could be efficiently converted in covalent bonds as shown by the groups of Otto[39] and Ashkenasy,[37] Tommaso Fraccia et al. have tested this

hypothesis using a DNA oligomer in a liquid crystal state. The results of this study show that the addition of a condensing agent such as the carbodiimide EDC to a liquid crystal solution of DNA induces its ligation with a much higher efficiency if compared to the same reaction in an isotropic state.[40]

The reaction enhanced by the liquid crystal state is the first case, to our knowledge, of a non-enzymatic polymerization reaction in nucleic acids that is not enhanced by a complementary strand behaving as a template or an external source either templating or catalyzing the reaction, but it is templated through end-to-end interactions of the nucleic acid molecules themselves.

The reactivity enhancement is not the only interesting properties connected to the liquid crystallization of nucleic acids. Tommaso Bellini et al.[41] have shown that a mixture of completely random DNA sequences still retains its capability to undergo liquid crystallization. In particular, it was shown that a self-sorting mechanism drive the formation of long linear aggregates among molecules capable to interact through terminal overhangs, while molecules rich in defects or lacking sticky ends become excluded.

Moreover, in the past year we have shown that even triphosphorilated nucleosides can produce liquid crystals,[3] removing the lower size limit for the supramolecular organization of nucleic acids, even if the details modulating this phenomenon are not clear.

The extension of these results on RNA and a deeper comprehension of the mechanism behind the emergent properties of supramolecular ordering would make liquid crystals of nucleic acids an appealing *guiding hand* towards the formation of ribozymes, based uniquely on the very chemical-physical properties of nucleic acids themselves.

Results and Discussion

2.1 Enhancement of RNA ligation efficiency by LC supramolecular ordering

Data shown in section 2.1 of this work have been partially collected during my master thesis and reinforced and expanded during my PhD. Here these results will be briefly exposed for the sake of completeness and in order to allow the understanding of the results that will follow in the next sections.

2.1.1 Liquid cystal phases in pRNA samples

Given the stricking observation that the liquid crystalline state in short DNA oligonucleotides solutions does enhance their ligation efficiency,[40] we wanted to understand whether this effect could be extended to RNA in order to raise the relevance of this phenomenon in the prebiotic context, where RNA is considered a more ancient nucleic acid. We decided to work with two new molecules synthesized by NOXXON pharma:

12BE: a 12nt long self-complementary 5'-phosphate RNA oligomer of sequence CGCGAAUUCGCG. 12BE produces a blunt double helix capable to organize as linear aggregates only through stacking interaction. This molecule is made of L-RNA and thus cannot be degraded enzymatically, reducing the variables to be taken into account

in this study.

Stacked double helices organize as follows (first line has 5'-3' polarity):

C G C G A A U U C G C G | C G C G A A U U C G C G |
G C G C U U A A G C G C | G C G C U U A A G C G C |

6SE: a 6nt long partially self-complementary 5'-phosphate RNA oligomer of sequence GAUCGC. 6SE produces double helices with 4 paired bases and two overhanging nucleotides. These overhangs can interact to stabilize the end-to-end aggregation of these molecules.

Stacked double helices organize as follows (first line has 5'-3' polarity):

G A U C G C | G A U C G C | G A U C G C | G A U C G C |
| C G C U A G | C G C U A G | C G C U A G | C G C U A G |

For both molecules we defined a phase diagram as a function of temperature and RNA concentration, where every column in the plots shown in Fig. 2.1 corresponds to an independent microscope cell prepared at a different target concentration.
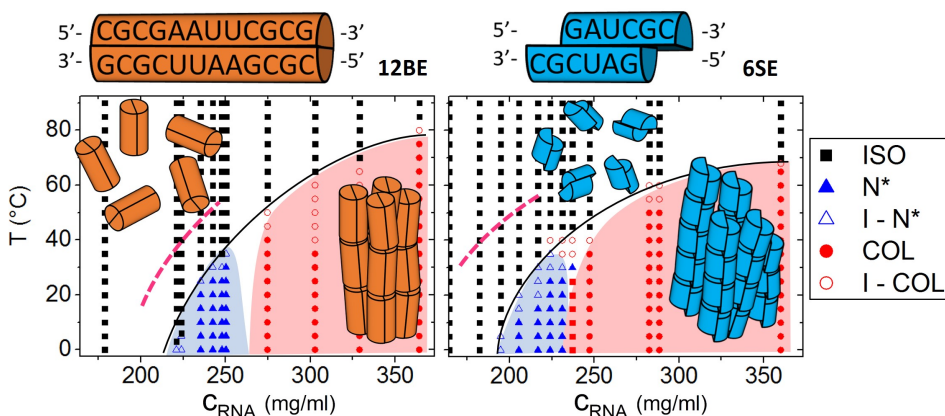
Fig. 2.1 Phase diagrams (T vs c_{RNA}) for 12BE (left) and 6SE (right). Molecules structure is sketched on top. Black line represents the phase boundary between isotropic and liquid crystalline phases. Pink dashed lines represent the phase boundaries after addition of 60mM $MgCl_2$, which stabilizes the phases.

The phase diagrams of 12BE and 6SE are remarkably similar. This might be surprising at first since double the elements should aggregate to overcome Onsager's limit in 6SE mixture. Nevertheless, the sticky interaction at the double helices termini is strong enough to compensate the reduced size. The addition of 60mM $MgCl_2$ to the samples stabilizes considerably the phase transition shifting the boundary at lower concentrations by roughly 25g/l. This observation underlines the critical role of ionic composition and multivalent cations, that are known to behave in a complex way modulating bridging between nucleic acid double helices.

2.1.2 PEG-pRNA phase separation

To run an efficient ligation reaction we make use of a large excess of the condensing agent *1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide* (EDC) which has the downside of disrupting the liquid crystalline textures at concentrations above $\approx 50mM$.

To overcome this issue we exploit the phase separation induced by the addition of the chemically inert and flexible polymer PEG (polyethylene

glycol, average $MW \approx 8000u$) to short nucleic acid oligomers mixtures (see Chapter 1.1.2.2 for more details).
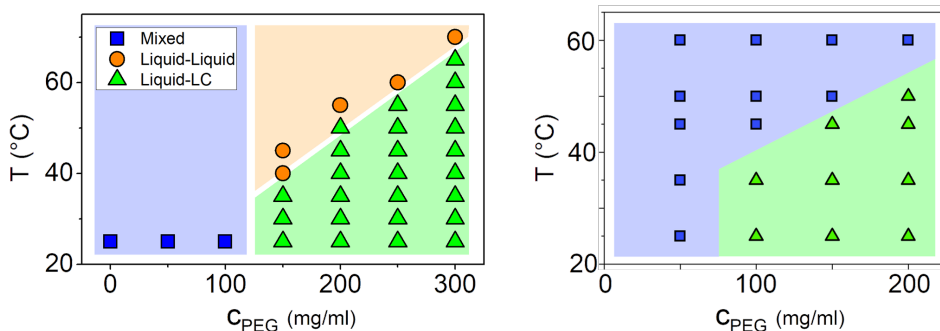
Fig. 2.2 Phase diagrams (T vs c_{PEG8k}) for 10mM 12BE, 60mM MgCl$_2$ (left) and 50mM 6SE, 60mM MgCl$_2$ (right).

As more and more PEG is added to a solution of either 12BE or 6SE the mixture transitions from being homogeneous (from now on defined as *Mixed* state and marked with blue color) to phase separated (from now on defined as *iso-iso* of *liquid-liquid* state and marked with orange color). In this phase separated state we can observe small droplets of highly concentrated RNA dispersed into a bulk of PEG. PEG behaves as an osmolyte and as more is added to the solution the droplets shrink due to the increase in osmotic pressure and the content increases in concentration and can transition to a liquid crystal state (from now defined as *LC-iso* or *liquid-LC* state and marked with green color).

For 12BE at 40°C we observe first the phase separation of the system at c_{PEG8k} = 150g/l and then the transition to liquid crystal state of the RNA contained in the droplets at $c_{PEG8k} \geq$ 200g/l. For 6SE regardless of the temperature this two-steps transition is never observed and the system goes directly from mixed with PEG to phase separated with liquid crystalline domains. A more detailed study of the multi-parameter variable space (ionic composition, ions concentration, PEG size, PEG concentration, temperature...) might allow to find a trajectory where an iso-iso state is found in 6SE.

For the purposes of our study 12BE is a better choice to decompose the effect of phase separation from the effect of supramolecular ordering, and will be studied in more details.

2.1.3 Ligation reactions of pRNA with EDC

To study the effect of phase separation and liquid crystal transition on the chemical reactivity of RNA, ligation experiments have been performed in microscope cells with a sample of 1μl of final volume, prepared with either 10mM 12BE, 60mM $MgCl_2$, 10mM Hepes pH 7.55 and 1.8M EDC or 50mM 6SE, 60mM $MgCl_2$, 10mM Hepes pH 7.55 and 1.25M EDC. The samples studied in Fig. 2.3 have been reacting for 48h and monitored over time through polarized optical microscopy to confirm that the phase separation was stable and that liquid crystals were not disrupted. Reactions were stopped adding tenfold volume of 50mM ethanolamine to reaction mixtures and products have been separated thought denaturing PAGE to quantify the reaction yields as described in details in materials and methods.
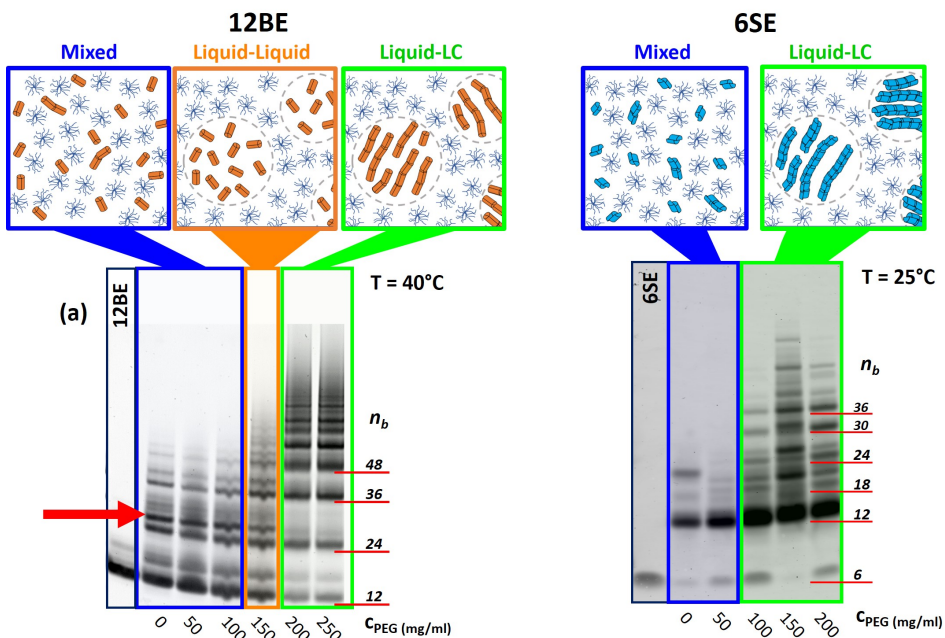
Fig. 2.3 On the left denaturing PAGE 15% of reaction products of 12BE after 48h of reaction at 40°C. On the right denaturing PAGE 20% of reaction products of 6SE after 48h of reaction at 25°C. Insets on top depict the physical state of the pRNA (sketched as cylinders) and PEG (sketched as blue clews) mixtures varying the concentration of the latter.

Interestingly, temperature has a non trivial effect on the ligation yield for 12BE as shown in Fig.2.4. While in pDNA[40] a great enhancement of reaction yield is observed even at low temperature, this is no longer true for pRNA, where a discontinuity can still be observed between the mixed state and the LC-iso state, but having a much smaller entity with respect to the one observed at 40°C.
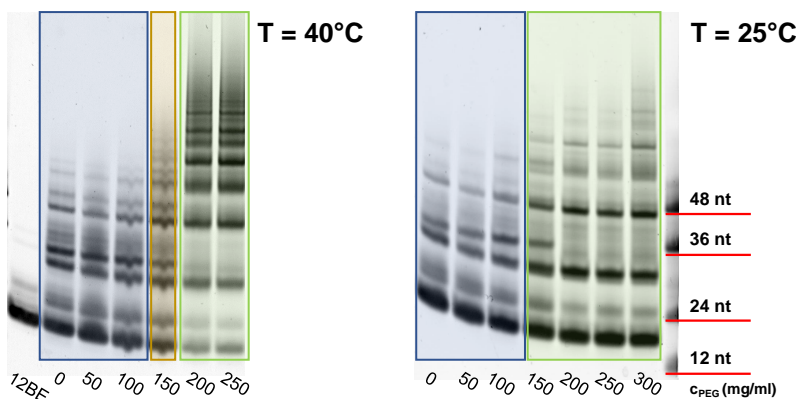
Fig. 2.4 Comparison of reaction products for 12BE at 40°C and 25°C.

From these gels the yields can be extracted as detailed in Chapter 4.5. The results of the analysis shown in Fig. 2.5 support the presence of a modest discontinuity in the ligation yield at room temperature, while the enhancement is much more marked at 40°C. The lines are linear fits to the data to better visualize the discontinuities; black lines at 25°C, red lines at 40°C. Each point corresponds to an independent experiment.
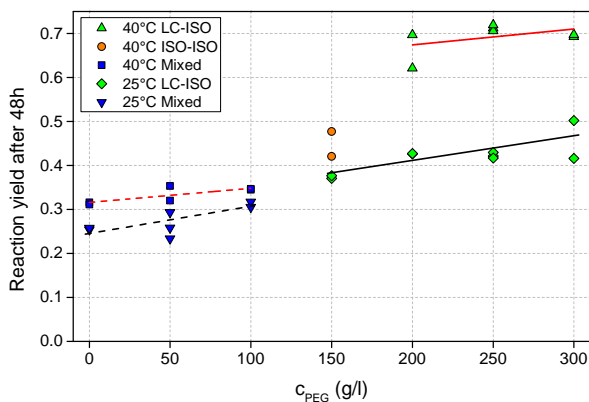
Fig. 2.5 Yields measured in EDC-mediated ligation of 12BE.

While tempting to increase further the temperature to obtain a higher

effect from the LC just based on this observation, increasing the temperature above 40°C induces such a high hydrolysis that after 2h almost no band is detectable in the PAGE experiments. In Fig. 2.6 the maximum yield measured at different temperatures is plotted. For the reactions occurring at 25°C and 40°C the data for yield at 48h are plotted. For the reaction occurring at 60°C the yield after 2h is plotted.
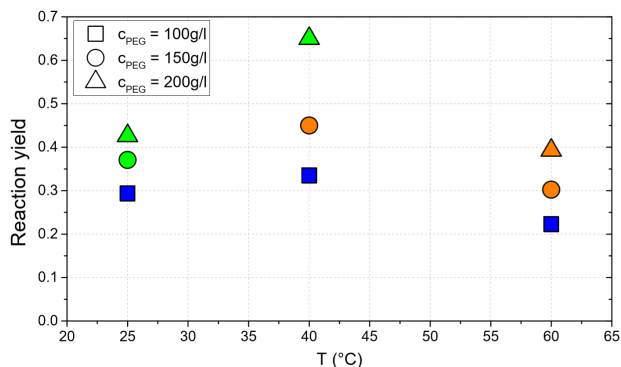
Fig. 2.6 Dependence of maximum reaction yield on temperature for 12BE.

At higher PEG concentration we consistently measure higher reaction yields, as expected from a simple concentration effect, while at fixed PEG concentration we see a non-monotonous yield as a function of temperature. This effect is due to a strong hydrolysis pushing the system towards small nucleic acid sizes. At 60°C this phenomenon is so fast that no electrophoretic band can be detected after two hours of reaction.

At the temperature resolution of this study we can say that at 40°C both the yields measured and the LC enhancement are maximized, making it a sweet spot for our study.

2.2 Quantification of pRNA concentration in phase separated domains

Since concentration is a fundamental parameter in determining a reaction rate, in order to decompose its effect it is necessary to quantify the concentration of nucleic acids in the RNA-rich domains and in the PEG-rich reservoir. Given a set of solutions with varying RNA concentration at a given constant PEG concentration, it is found that phase separation occurs when RNA concentration c_{RNA} is higher than a concentration threshold c_{RNA}^T defined as the maximum solubility of RNA in the PEG mixtures. We thus assume that as c_{RNA} gets above the threshold the concentration in the PEG-rich phase stays $c_{out} = c_{RNA}^T$, and all the additional material increases the volume fraction occupied by the RNA-rich phase.

It must hold true that:

$$c_{RNA} \cdot V_{tot} = c_{in} \cdot V_{in} + c_{out} \cdot V_{out} \tag{2.1a}$$

$$c_{RNA} = c_{in} \cdot \phi_{in} + c_{out} \cdot \phi_{out} \tag{2.1b}$$

$$c_{RNA} = c_{in} \cdot \phi_{in} + c_{RNA}^T \cdot (1 - \phi_{in}) \tag{2.1c}$$

$$c_{RNA} = c_{in} \cdot \phi_{in} + c_{RNA}^T - c_{RNA}^T \cdot \phi_{in} \tag{2.1d}$$

$$c_{RNA} = \phi_{in}(c_{in} \cdot c_{RNA}^T) + c_{RNA}^T \tag{2.1e}$$

We can then rewrite equation 2.1e as explicit in c_{RNA}:

$$\phi_{in} = \frac{1}{c_{in} - c_{RNA}^T} \cdot c_{RNA} - \frac{c_{RNA}^T}{cin - c_{RNA}^T} \tag{2.2}$$

Given that we can experimentally measure ϕ_{in} and that c_{RNA} is given

by sample preparation, it is possible to determine the following parameters:

$$m = \frac{1}{c_{in} - c_{RNA}^{T}} \tag{2.3a}$$

$$q = -\frac{c_{RNA}^{T}}{c_{in} - c_{RNA}^{T}} = -c_{RNA}^{T} \cdot m \tag{2.3b}$$

From which we can obtain:

$$c_{in} = \frac{1 - q}{m} \tag{2.4a}$$

$$c_{RNA}^{T} = -\frac{q}{m} \tag{2.4b}$$

For our study we have focused on samples at $c_{PEG8k} = 150g/l$ and $c_{PEG8k} = 200g/l$.

In Fig. 2.7 it is shown the effect of increasing 12BE concentration in a mixture with c_{PEG8k} = 150g/l and MgCl$_2$ 60mM. In the first micrograph on the left, the sample is right above $c_{RNA}{}^{T}$, and as more pRNA is added, the more the RNA-rich phase increases in occupancy, while showing the same LC texture, confirming this way that its concentration remains constant.
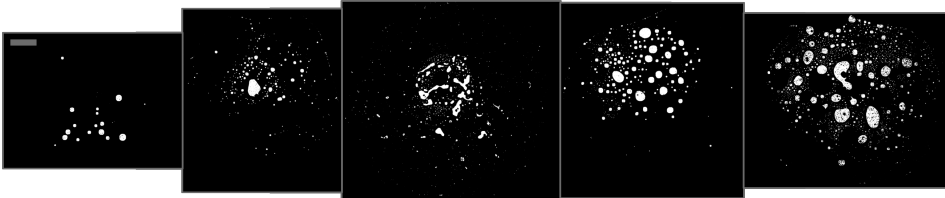
Fig. 2.7 Processed maps of 12BE samples birefringence in mixtures with PEG8k 150g/l. From left c_{RNA} is equal to 13g/l, 19g/l, 25g/l, 28g/l and 38g/l. Bar length is equal to $500\mu m$. As pRNA concentration increases, the volume fraction of the samples filled by liquid crystals grows.

As already discussed in the Chap. 1.1.2.2, the entropy-driven phase separation emerging in RNA-PEG mixtures is due to their difference in flexibility. Previous studies from Giuliano Zanchetta[42] have shown that

the same phase separation occurs even in mixtures of DNA double helices and DNA single stranded oligonucleotides. This is particularly interesting when dealing with RNA, since functional RNA molecules are generally flexible polymers capable to fold in the three dimensional space, and they could be expelled by the liquid crystals of surrounding rigid double helices of RNA. The results of the study of the phase separation of 12BE with PEG8k and with a 11nt long single stranded RNA ($c_{ssRNA} = 50g/l \approx 10mM$) are summarized below:
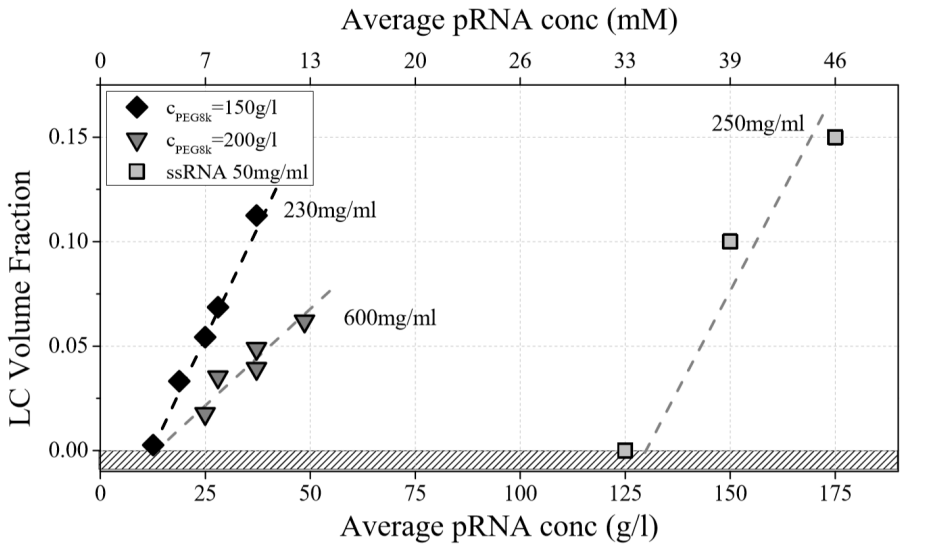
Fig. 2.8 Scatter plot for ϕ_{in} for 12BE samples. As expected from theoretical considerations, in a phase separated mixture of double stranded pRNA with either PEG8k or ssRNA (single-stranded RNA), as pRNA concentration increases, the volume fraction occupied by liquid crystals grows. The weaker the process, the greater the osmotic pressure produced by the flexible polymer.

From this analysis we could obtain values for $c_{in}^{PEG8k150g/l} = 230g/l$, $c_{in}^{PEG8k200g/l} = 600g/l$ and $c_{out}^{PEG8k} = 13g/l$ for both 150g/l and 200g/l. The study of the ssRNA gave as results $c_{in}^{ssRNA} = 250g/l$ and $c_{out}^{ssRNA} = 125g/l$.

2.3 Solutes partition unequally in RNA-rich and PEG-rich phases phases

Among our assumptions to model EDC reaction in the phase separated system is that EDC concentration is spatially uniform in the sample over time. We investigated whether relatively small ($MW \approx 300 - 500 g/mol$) fluorescent molecules could equally distribute between the highly concentrated pRNA domains and the PEG outside. What we could find is that the fluorescent signal of Fluorescein inside the pRNA domains is much lower than in the surrounding PEG-rich solution.

A study of Fluorescein signal as a function of pRNA concentration (pH 7.5) shows that the fluorescent intensity is actually quenched by the presence of nucleic acids. Changing the concentration from $c_{out} \approx 13 g/l$ to $c_{in} \approx 230 g/l$ we should expect the signal to be reduced by $\approx 65\%$. In our system we see a reduction of $\approx 90\%$, meaning that while a great part of the big fluorescence difference in the two phases should reasonably be due to quenching, it is not enough to justify this effect, which can be indeed due to differential partitioning.
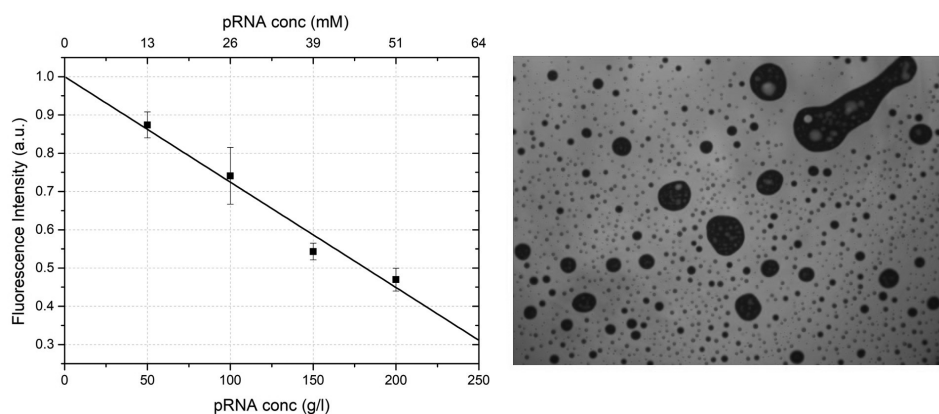
Fig. 2.9 On the left fluorescence intensity of Fluorescein as a function of pRNA. On the right micrograph of the fluorescence emission of Fluorescein in a phase separated sample of pRNA and PEG8K 150g/l.

To determine whether the difference in signal could reflect a difference in concentration regardless of Fluorescein chemical details, which is a negatively charged molecule at neutral pH, we reproduced the experiment using Rhodamine B and Safranin, both positively charged.
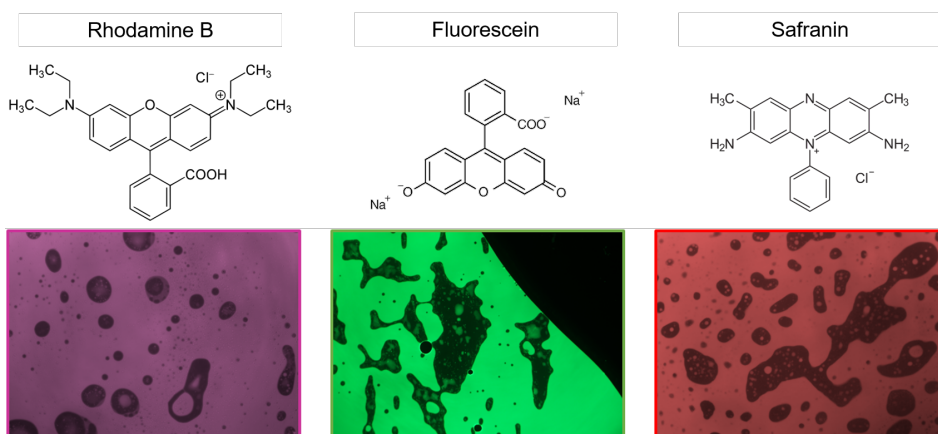
Fig. 2.10 False color micrographs of fluorescence emission of Rhodamine B, Fluorescein and Safranin in samples of 12BE in PEG8k 150g/l. Dark areas correspond to RNA-rich droplets.

Also in these cases a remarkable difference in fluorescence intensity can be observed, reinforcing the interpretation of a differential solubility of the molecules for the two phases. Interestingly, the difference is intensity is more modest with the positively charged molecules, as expected. Given the fluorescence quenching observed in this study, in the future different approaches should be exploited to study the partitioning of solutes.

2.4   EDC diffusion is not limiting the reaction yield

Dealing with a system with incredibly high concentrations and with mesophases, the samples reach a very high viscosity that could impair the diffusion of the reacting species. Since the EDC consumption occurs much faster in the RNA-rich domains behaving as microreactors, we might approach a diffusion-limited regime in which EDC is depleted and limits the reaction which occurs on a timescale faster than its diffusion from the PEG-rich phase. In order to rule out whether the reactions in our samples could be diffusion-limited for EDC, we investigated the diffusion time of a small molecule (Fluorescein) in and out of the liquid crystal domains.

A series of FRAP experiments have been performed as described in Chap. 4.8 to obtain diffusion coefficients of Fluorescein mixed in different samples. The viscosities measured this way, using an hydrodinamic radius $R_h = 0.45nm$ for Fluorescein are as follows:

Table 2.1 Viscosities measured for 12BE and PEG8k system

pRNA Sample	Viscosity $(mPa \cdot s)$
LC 230g/l + MgCl$_2$ 60mM	48.3
ISO 200g/l + MgCl$_2$ 60mM	8.2
ISO 150g/l + MgCl$_2$ 60mM	4.8
ISO 100g/l + MgCl$_2$ 60mM	3.3
ISO 50g/l + MgCl$_2$ 60mM	2.7
PEG8k Sample	**Viscosity** $(mPa \cdot s)$
400g/l	6.9
200g/l + MgCl$_2$ 60mM	9.3
200g/l	4.1
150g/l + MgCl$_2$ 60mM	2.1

Assuming EDC diffuses on a comparable timescale given its similar size with Fluorescein, we can estimate a diffusion coefficient equal to

$\approx 15\frac{\mu m^2}{s}$, which corresponds to a much faster process than the reaction characteristic timescales. Since the size of RNA-rich domains is in the order of tens of micrometers, only few seconds are necessary to equilibrate EDC concentration with the surrounding PEG-rich phase. It follows that EDC diffusion through the two phases is not limiting the reaction.

2.5 A model for EDC-mediated pRNA condensation

To drive the chemical ligation of 12BE and 6SE we used *1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide* (EDC), a chemical compound belonging to the class of the carbodiimides, whose reactivity has been widely studied by Shabarova group.[43] The compound has a strong UV absorbance (red line in Fig.1) and over time it is converted to *1-Ethyl-3-(3-dimethyl-aminopropyl)urea* (EDU), with a clearly different UV spectrum (green line in Fig.1).[44] In aqueous solution the evolution of the measured spectra can be described as the linear superposition of the spectra of the two compounds.
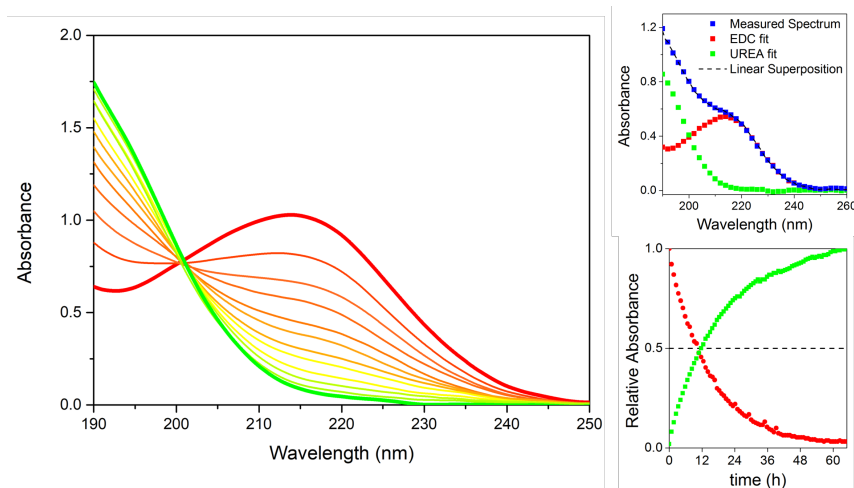
Fig. 2.11 EDC UV spectrum evolution over time.

The simple decay of EDC signal measured through UV absorbance can be modeled as a single exponential $e^{-k_h t}$, where k_h is the kinetic constant of EDC hydrolysis. The parameter measured this way at 40°C, pH 7.3 and $[MgCl_2] = 60$mM is $\approx 0.033 h^{-1}$.

When EDC is used with a phosphate moiety it activates it producing an O-acilisourea intermediate, which quickly becomes hydrolyzed by water or reacts with an hydoxyl group to produce a natural phosphodiester bond.
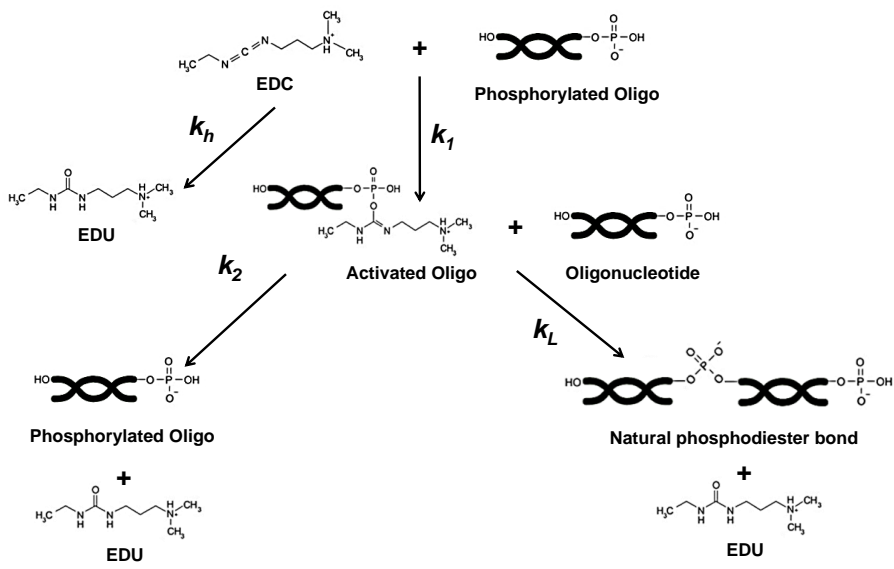
Fig. 2.12 Reaction scheme for EDC.

In more details, the reaction mechanism of EDC with a phosphoester group can be described as follows, clarify why the reaction is so sensitive to pH:
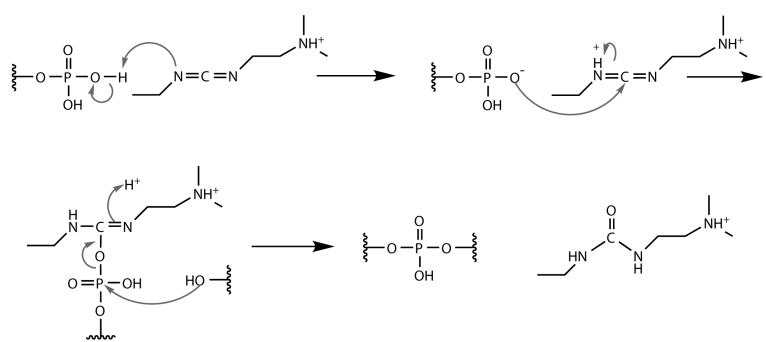
Fig. 2.13 EDC mechanism of reaction.

To characterize the k_1 parameter while sticking with the simple UV absorbance method, we could not use a proper ribonucleotide, given the high absorbance in the ultraviolet range of the nitrogenous bases as described in

Chapter 1.1.2.1.

To overcome this issue, we added to our EDC solutions a small amount of 5'-phosphate ribose to mimic the reactivity of a nucleotide while missing the nitrogenous base. The addition of a sugar-phosphate moiety does indeed speed-up the reaction as expected. This phenomenon can be modeled by adding a second kinetic parameter for the activation of the phosphate group.

$$c \cdot e^{-(k_h + [P]k_1)t} \tag{2.5}$$

Where c is EDC initial concentration and $[P]$ is ribose phosphate initial concentration.

The resulting estimate of k_1 is reliable under the following assumptions:

i The reaction intermediate can be neglected. This point is supported by the observation that EDC and EDU spectra evolve with the same rate and the measured spectra are well fitted by a linear superposition of those of EDC and EDU alone. This indicates that the intermediate species either has no UV absorbance or has a negligible concentration. The most simple explanation in according to literature is that its decay rate k_2 is much larger than k_1. This also implies that [P] is basically constant.

ii A negligible amount of phosphodiester bonds are formed during the reaction. This is a reasonable assumption that takes into account the low yield of phosphodiester bond formation in our ligation reactions (as determined through PAGE analysis), and the low concentration of P used in this set of experiments, ranging from 0.5 mM to 10 mM, being ligation efficiency proportionally to $[P]^2$. With this approach, we could determine a $k_1 \approx 0.020 h^{-1} mM^{-1}$.

To finally estimate the two missing parameters, k_2 and k_L we have to solve

the following set of differential equations:

$$\frac{d[EDC]}{dt} = -[EDC]k_h - [EDC][P]k_1 \tag{2.6a}$$

$$\frac{d[P]}{dt} = -[EDC][P]k_1 + [P^*][H_2O]k_2 \tag{2.6b}$$

$$\frac{d[P^*]}{dt} = [P]k_1 - [P^*][H_2O]k_2 - [P^*]([P^*] + [P])k_L \tag{2.6c}$$

$$\frac{d[S]}{dt} = ([P^*] + [P])k_L \tag{2.6d}$$

$$\frac{d[EDU]}{dt} = k_h + [P^*][H_2O]k_2 + [P^*]([P^*] + [P])k_L \tag{2.6e}$$

Where [EDC], [EDU], [P], [P*], and [S] are the molar concentrations of EDC, EDU, 5-phosphate RNA oligomers, activated 5-phosphates RNA oligomers and newly formed phosphodiester bonds, respectively. This system of equations, combined with the conservation constraints between related species such as [P], [P*] and [S], can be solved to obtain the evolution of the concentration of each species over time. Since no analytical solution can be given, we had to solve the system of linear differential equation numerically using the ODE45 function included in MATLAB. We solved the system fitting our experimental data through an implemented fminsolve function to search for the best k_2 to k_L ratio (R = k2 / kL) that could reproduce our experimental data in the mixed state (see Fig. 2.14 left panel). Using this approach we determined $R \approx 1.1110^{-2}$. Since the reaction yield is ultimately determined by a balance of the tendency of the system to react versus water or versus a ribose hydroxyl group, it becomes convenient to define this tendency using a pure number, defined as a ratio between the kinetic parameters k_2 and k_L. In order to do so, k_2 has to be treated as a second order reaction with water, where [H2O], which is defined as a reactant in the equations 2.6, is set at a constant 55M.

This system of equations aims at the description of the reaction in isotropic state but cannot take into account the liquid-liquid phase sepa-

ration occurring when PEG concentration is increased above 150g/l. To model the phase-separated system we must implement a new set of equations as follows:

$$\frac{d[EDC]}{dt} = -[EDC]k_h - [EDC]([P_{in}]F_{in} + [P_{out}]F_{out})k_1 \qquad (2.7a)$$

$$\frac{d[P_{out}]}{dt} = -[EDC][P_{out}]k_1 + [P^*_{out}][H_2O]k_2 \qquad (2.7b)$$

$$\frac{d[P_{in}]}{dt} = -[EDC][P_{in}]k_1 + [P^*_{in}][H_2O]k_2 \qquad (2.7c)$$

$$\frac{d[P^*]}{dt} = [EDC]([P_{in}]F_{in} + [P_{out}]F_{out})k_1 - ([P^*_{in}] + [P^*_{out}])[H_2O]k_2 +$$
$$- [P^*_{in}]([P^*_{in}] + [P_{in}]F_{in})k_L - [P^*_{out}]([P^*_{out}] + [P_{out}]F_{out})k_L$$
$$(2.7d)$$

$$\frac{d[S]}{dt} = [P^*_{in}]([P^*_{in}] + [P_{in}]F_{in}) \cdot k_L + [P^*_{out}]([P^*_{out}] + [P_{out}]F_{out}) \cdot k_L$$
$$(2.7e)$$

$$\frac{d[EDU]}{dt} = [EDC]k_h + ([P^*_{in}] + [P^*_{out}])[H_2O] \cdot k_2 +$$
$$[P^*_{in}]([P^*_{in}] + [P_{in}]F_{in}) \cdot k_L + [P^*_{out}]([P^*_{out}] + [P_{out}]F_{out}) \cdot k_L$$
$$(2.7f)$$

In this system of linear differential equations, the parameters $[P_{in}]$, $[P_{out}]$, $[P^*_{in}]$, and $[P^*_{out}]$ are the concentrations of terminal phosphates and activated terminal phosphates inside the RNA-rich domains and outside, respectively. To model the reaction it is also necessary to introduce two new parameters F_{in} and F_{out} for the volume fractions of the two phases: $F_{in} = \frac{V_{in}}{V_{tot}}$, $F_{out} = \frac{V_{out}}{V_{tot}}$.

The concentration of RNA in the two phases and their volume fractions have been determined as in Chapter 2.2. In order for the model to be reliable we have to assume that RNA partitioning between the two phases is stable over the reaction, i.e. that $[P_{in}] + [P^*_{in}]$ and $[P_{out}] + [P^*_{out}]$ are constant in time, supported by the fact that the morphology of the samples

remains unaltered if the temperature of the system does not exceed too high temperatures. When solving this system of equation with the values of k_h, k_1, k_2, and k_L determined in the mixed state, the reaction yield predicted greatly overestimates the one experimentally measured from PAGE experiments in isotropic-isotropic phase separated state (see Fig.2.14 central panel dashed line). To solve this issue and have a complete model we should include (i) the effect on the kinetic parameters of the larger viscosity of RNA-rich phases, measured in Chapter 2.4, and (ii) the possible unbalanced partitioning of EDC in the two phases due to reduced solubility in the RNA-rich phase. Both these factors have the effect of slowing down the ligation yield in the RNA-rich domains. For sake of simplicity, we decided to take into account these effects by modifying the rates controlling the reaction of RNA with EDC (k_1) and the rate controlling the ligation reaction (k_L) within RNA-rich domains. We reasonably expect k_1 and k_L to slow down at least proportionally to the viscosity, which is at least proportional to RNA concentration, we could modify the kinetic parameters within the RNA-rich domains, where these effects become are relevant, by replacing k_1 and k_L with $k_1 = \frac{F}{c_{RNA}}k1$; $k_L = \frac{F}{c_{RNA}}k_L$, where F is a constant parameter.

To determine parameter F we numerically solved the system of equations 2.7 to reproduce the ligation kinetics measured in liquid-liquid coexistence at $c_{PEG} = 150g/l$, which gives a parameter F = 30 mM, (see Fig. 2.14 central panel dotted line). This new parameter contains the information for the dependence of k_1 and k_L on RNA concentration and can thus be used to predict the expected yield in samples with LC-liquid coexistence, dissecting the contribution of liquid crystal ordering and concentration alone. Solving this system of equations using the concentration of RNA and volume fractions determined in the LC-liquid coexistence, the resulting yield is much lower than the measured yield (see Fig. 2.14 right panel dotted line). It is important to stress that this underestimate would become way more dramatic if we used a more accurate dependence of ki-

netic parameters on the measured viscosity. Given that the experimentally measured yield is much larger than the one predicted by the model under this set of assumptions, an enhancement of ligation yield given by LC order can be invoked and characterized.
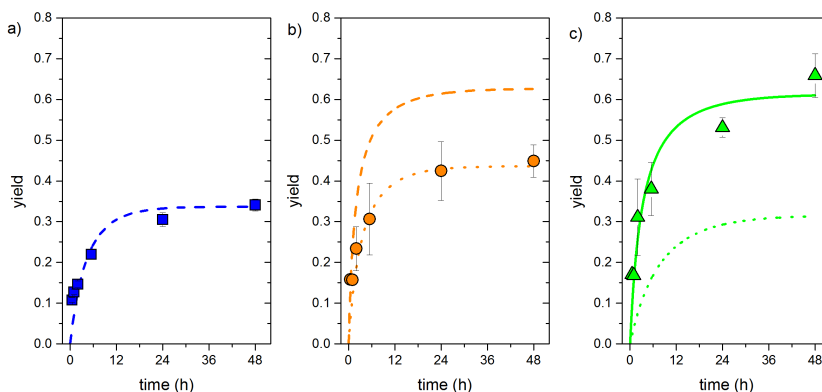
Fig. 2.14 Scatter plot for reaction yield in isotropic state (left), isotropic-isotropic phase separation (center) and LC-liquid state (right). Lines are predictions obtained by numerically solving the systems of differential equations reported in the text. For the isotropic-isotropic phase separation a prediction is made both when the growth in viscosity is accounted for (dotted-line) or not (dashed-line). For the LC-liquid state the dotted line is the predicted yield using the same kinetic parameters determined in the previous cases, while the continuous line is the yield predicted with an increase in the kinetic constant k_L.

The complexity of modeling a phase-separated system with such high concentrations forced us to keep in account three competing effects:

i) higher RNA concentrations increase activation and ligation reaction speeds;

ii) higher RNA concentrations increase the viscosity of the system, reducing the reaction rates[45] and making hydrolysis or the active intermediate more relevant;

iii) higher RNA concentration in the phase separated domains reduce the solubility of other molecular species, which could limit EDC re-

action rate with RNA.

In this work we aimed at capturing the effects of (ii) and (iii) introducing a parameter F, whose role is to take into account the reduction of reactivity of RNA towards EDC and activated RNA molecules towards other molecules.

By solving the system of equations 2.7 varying c_{in} while holding c_{out} for pRNA equal to 13g/l (as estimated in Chapter 2.2) and the volume fractions F_{in} equal to 0.10 to describe the phase separated system with c_{PEG}=150g/l and 0.04 to describe the system with c_{PEG}=200g/l we find a function describing the yield which is non-monotonic as shown in Fig. 2.15, where a maximum is present at a concentration which is actually lower than the ones of the actual RNA-rich phases, equal to 60mM in c_{PEG}=150g/l and 158mM in c_{PEG}=200g/l.
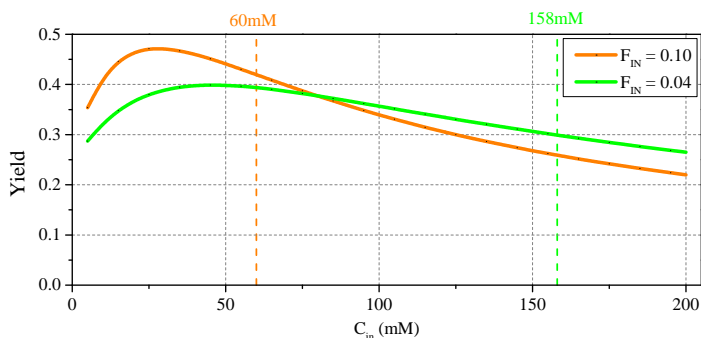
Fig. 2.15 Ligation yields predicted as a function of 12BE concentration in the pRNA-rich domains in the isotropic-isotropic phase-separated system (orange continuous line) at a fixed pRNA volume fraction equal to 0.10 or LC-liquid state (green continuous line) at a fixed pRNA volume fraction equal to 0.04. The slow-down of the kinetics due to the growth of viscosity determines an absolute maximum in reaction yield lower than 0.5, which is overcome by the LC ordering as discussed in the text.

Given this description of the system, one should not expect to measure a yield higher than ≈ 0.5 in an isotropic phase separated state. Nevertheless,

in the in the liquid crystalline domains in c_{PEG}=200g/l, we could measure a reaction yield of ≈ 0.7, much higher than expected, meaning that the liquid crystal state is enhancing the reactivity.

According to Arrhenius Law, to undergo a chemical reaction the reactants must have a Gibbs free energy of activation defined as:

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger} \tag{2.8}$$

Aside from the very well known mechanisms of chemical catalysis operated by enzymes to reduce the activation energy of a target chemical reaction, their structures is optimized to interact with the reactants and position them to reduce their *entropy of activation*, reducing the space of possible conformations that they can assume,[46] a mechanism that goes under the name of *proximity and orientation* of the substrates.

As the system transitions from isotropic to liquid crystal, the collective organization stabilizes long end-to-end aggregates and causes a loss of translational and rotational entropy due to the very nature of stacking interaction. Full-atom simulations by Aksimentiev group[47] support the notion that two DNA blunt double helices kept in close proximity by stacking interaction are not free to rotate one with the respect to the other when they are functionalized with a 5'-phosphate terminus. A preferential orientation exist in which the two helices keep in close proximity the 3'-hydroxyl group of a DNA molecule with the 5'-phosphate group of the adjacent one, in an optimal configuration for the formation of a phosphodiester bond.

Following this direction, we can reproduce the unexpectedly high yield of reaction in the liquid crystal state by replacing the kinetic parameter for the reaction occurring the RNA-rich domains k_L with $k_{LC} = 6k_L$ to better approximate the experimental data (see continuous line in right panel Fig. 2.14).

From the enhancement of the kinetic parameter we can calculate a variation in the activation energy towards the transition state $\Delta\Delta G^{\ddagger} =$

$RTln\frac{k_{LC}}{k_L}$, equal to $1.1\frac{kcal}{mol}$, a modest value if compared to enzymes that also exploit a reduction of the *enthalpy of activation*, or the enhancement of intramolecular reactions in organic chemistry, but in agreement with the energy magnitude in play namely stacking interaction -.

Another result of this analysis is the possibility of decomposing the yield of the reaction in a phase-separated system as the sum of the yields in the two separate compartments. The analysis of the denaturing PAGE for the products of a reaction in a mixed isotropic state directly gives access to a yield and an average product size reflecting the behavior of the whole solution. For a phase separated system what we measure is the weighted sum of the reaction yields in the two compartments, where according to the previously established nomenclature S_{in} and S_{out} are the reaction products in the two compartments of volume fractions V_{in} and V_{out} and pRNA concentrations P_{in} and P_{out}.

For such a system, the yield quantified in our PAGE experiment is defined as:

$$yield_{measured} = V_{in}\frac{[S_{in}]}{[P_{in}]} + V_{out}\frac{[S_{out}]}{[P_{out}]} \qquad (2.9)$$

Through our analysis we can give estimates for the yields in the PEG-rich and pRNA-rich phases for the mixtures with c_{PEG}=150g/l and c_{PEG}=200g/l, where the colored symbols in the plot below correspond to the averagely weighted yield and the black and white symbol correspond to the average length of the products given the yield calculated in the two phases:
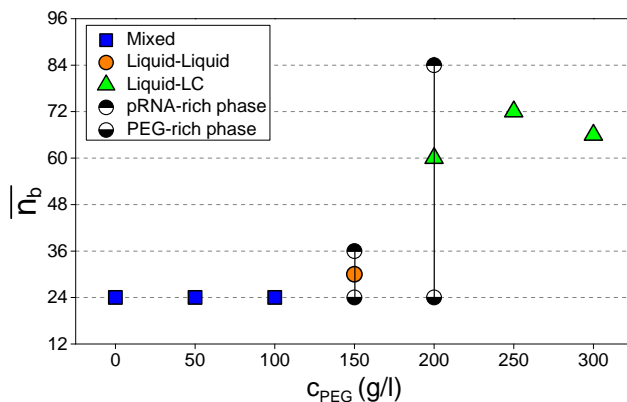
Fig. 2.16 Estimate of reaction yields from denaturing PAGE and kinetic analysis.

Our data point to a remarkable contribution to the global yield given by the liquid crystalline domains, that behave as microreators consuming EDC and generating products of a length which is 7 times bigger than the starting dodecamer, much larger than the average length of the products in the surrounding isotropic phase, equal to 2 times the starting dodecamer.

This is the first attempt at a mechanistic description for the enhancement of ligation yield in a nucleic acid liquid crystal. Even though we put a lot of effort to give a quantitative description for the reaction, we had to deal with an extremely complex and unusual system, that limited the potential of this work. In particular, a more exhaustive and robust modeling might come simplifying the system:

i We used EDC to drive the ligation reaction. The use of reactive species such as polyphosphorylated oligonucleotides or imidazole-phosphates derivatives would reduce the reaction scheme to a much simpler parallel competition between hydrolysis and ligation;

ii In order to use a great amount of EDC we had to exploit the phase separation induced by PEG. As a result of this choice we ended up with two coexisting phases where the concentrations of reactants can

differ by order of magnitudes. The use of a more efficient or tolerated condensing agent or pre-activated species would allow to study the ligation reaction without PEG and thus with a much better control on reactants concentrations.

iii We have measured the yields of reaction using a Flory model for simple polymerization (see Chapter 4.6), which has been developed and is thus well-suited to describe linear polymerization. As discussed in more details in the following Chapters 2.6 and 2.7 this is not what is happening when the reaction occurs in the isotropic state, thus making our quantification the more imprecise the less linear products are present.

2.6 Identification of circular products in 12BE

The transition to a liquid crystalline state corresponds to a symmetry breaking of the system, which is no longer isotropic and symmetric in every direction. Each molecule and nucleic acid aggregate diffuses and rotate in a preferential direction in space, and this affects the chemical reactivity of the system, as described above, but can also determine a different outcome of the chemical reaction.

When the EDC-reaction products in a liquid crystal state are separated in a denaturing PAGE, the electrophoretic bands are separated with the expected spacing for a discrete polymerization reaction. This is not true when the same is done for products of a reaction occurred in an isotropic state. For such a reaction, an unexpectedly greater number of bands is shown in denaturing PAGE experiments. Furthermore, not only the bands are more than expected, they are exactly double the expected number.

The most simple explanation for this discrepancy is that the surplus products are circularized polymers, an well-known issue linked to nucleic acid polymerization in the context of the Origin of Life.[48] This hypothesis is supported by the fact that these extra bands have lower electrophoretic mobility than the ones also present in the liquid crystal products.

To check whether this was true it is unfortunately not possible to directly image them through electron microscopy due to the limited size of the more abundant circular products, in the order of $\approx 10nm$ of linear length and $\approx 1nm$ of diameter. Indirect evaluation of their topology was obtained through 2D denaturing PAGE of 12BE reaction products, a widely used method to highlight nucleic acid molecules having unusual topologies.[49,50]

In Fig. 2.17a it is shown a PAGE 15% of reaction products as a function of PEG concentrations, where blue lanes correspond to products of a reaction occurring in mixed isotropic state, orange lanes correspond to a phase separated isotropic-isotropic state and the green lanes correspond to

a phase separated LC-isotropic state. The inlet on the left show our proposed topologies for the molecules resolved by the gel. In Fig. 2.17b we estimated the relative intensities of the lower and upper band pointed by the arrows in panel a. In Fig. 2.17c the lane chosen from first dimensional gel to be resolved in the second dimension is shown on the top of the second dimensional run (panel d). Molecules that lie on the diagonal share the same topology. Circular molecules are expected to move more slowly than linear nucleic acids and to lie above the diagonal.
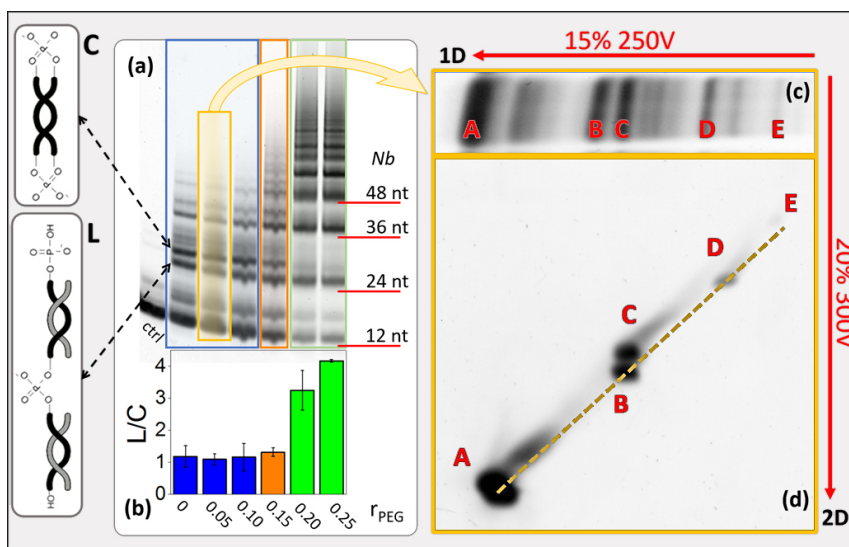
Fig. 2.17 2D PAGE to highlight molecules with circular topologies in a reaction occuring in an isotropic state ($c_{PEG} = 50g/l$).

This experiment shows that the band the we interpret as a 24nt circular product (C) is indeed shifted with respected to the other bands at low molecular weights, corresponding to a 12nt linear products (A), a 24nt linear product (B) and a 36nt linear product (D).

Interestingly, this phenomenon was not observed for ligation with the analogous DNA sequence in isotropic state.[40] This could be due to the slightly different reaction conditions, such as a different magnesium and EDC concentrations, or could be intrinsic to the molecule nature.

2.7 Identification of circular products in 6SE

Being 12BE a left-handed RNA molecule (synthesized using L-ribose), we could not treat it enzymatically to test further its topology, since enzymes have evolved to process biological RNA molecules made with D-ribose and are thus completely ineffective on 12BE. Since also in 6SE PAGE we can detect more bands than expected, we decided to treat the products of reaction of 6SE with Snake Venom Phosphodiesterase I (SVP) from *Crotalus adamanteus*.

This enzyme is known for its 3'-exonucleolytic and endonucleolytic activities, and it is known to be less effective on circular oligomers than linear oligomers.[51,52]
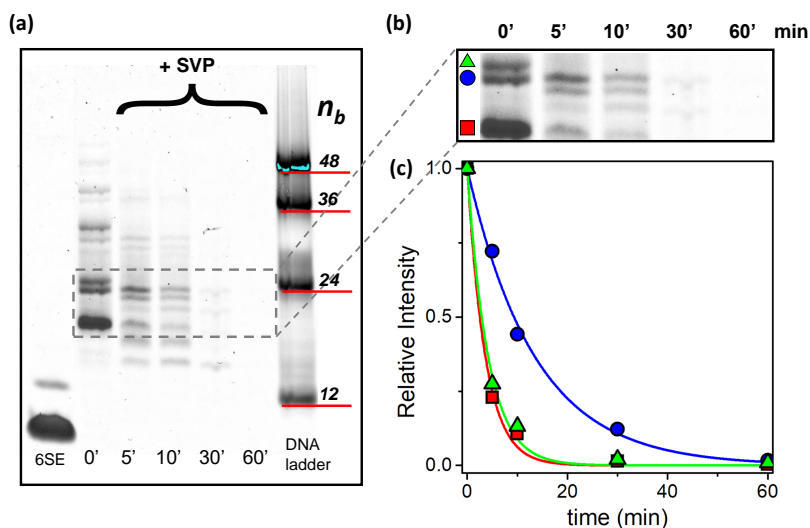
Fig. 2.18 Enzymatic digestion of 6SE reaction products with EDC. a) PAGE 20% denaturing gel of enzymatic digestions stopped at different times. b) highlight of low weight bands chosen for further analysis. c) Relative intensity of chosen bands over enzymatic digestion reaction.

A clear difference in efficiency is shown by the enzyme activity on different bands separated by the PAGE experiment, reinforcing our interpretation of reaction products being mixtures of linear and circular prod-

ucts. Interestingly, the liquid crystalline state of 6SE does not suppress the formation of this circular products, which are still well present in PAGE experiment.

2.8 Study of the relative stability of oligomers with different terminal modifications

Given that supramolecular ordering in the liquid crystal state requires the aggregation of molecules bearing negative charges (see Chapter 1.1.1) it is not of great surprise to find that electrostatic interaction, and thus counterions nature, play such a significant role in the formation of liquid crystals as shown for the stabilization of the phase transition for 12BE shown in Chap. 2.1 and the formation of liquid crystals of G-quadruplexes, as more extensively discussed in Chap. 2.10.

Aside from the importance of the net charge distribution over the molecule, we cannot neglect the fact that the formation of linear aggregates is mediated by end-to-end interactions. It is thus of major importance the evaluation of the impact on liquid crystals formation of different terminal modification adding more and more negative charges and functionalities to the molecules. We studied the phase diagrams for three molecules designed with the self-complementary Dickerson-Drew sequence CGCGAATTCGCG:

DD: DD molecule with 5'-hydoxyl terminus.
pDD: DD molecule with 5'-phosphate terminus.
3pDD: DD molecule with 5'-triphosphate terminus.

Knowing the pKa of these compounds we can predict their net charge at pH 7.0 as -11.0 for DD, -12.5 for pDD and -14.5 for 3pDD. Full atom simulations developed by Christopher Maffeo from Aksimentiev group[47] group (Fig. 2.19, unpublished) point at a reduction of the potential well for end-to-end interaction as more phosphate groups are added at the molecule terminus.
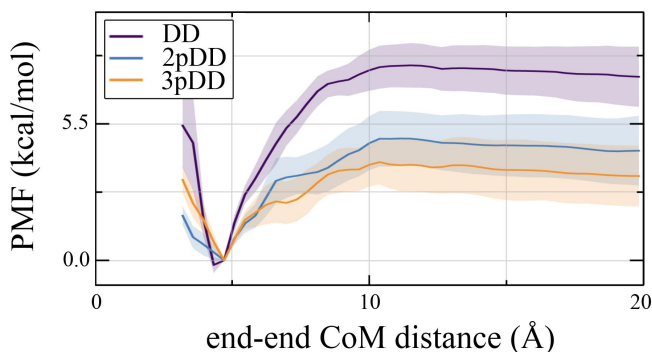
Fig. 2.19 Energy potential for end-to-end interaction in DD, 2pDD and 3pDD. Results obtained through a full-atom simulation by Christofer Maffeo (Aksimentiev group), unpublished.

Based on this prediction, we were expecting the supramolecular order to be suppressed in 3pDD. Nevertheless, we could find chiral nematic and columnar liquid crystalline phases in 3pDD samples, only greatly destabilized with respect to the ones observed in DD and pDD.

As shown in Fig.2.20a, 3pDD can organize in a cholesteric nematic phases with a pitch in the order of micrometers, that can be measured from the periodicity of maxima and minima of birefringence in the fingerprints-textures. The values obtained this way are plotted as a function of temperature. In Fig.2.20b a series of micrographs of cholesteric areas where the macrohelix is pointing towards the observer are shown. Since the birefringence is extinguished when the polarizers are uncrossed counterclockwise, the macrohelix of the N* phase of 3pDD is left-handed.[53]
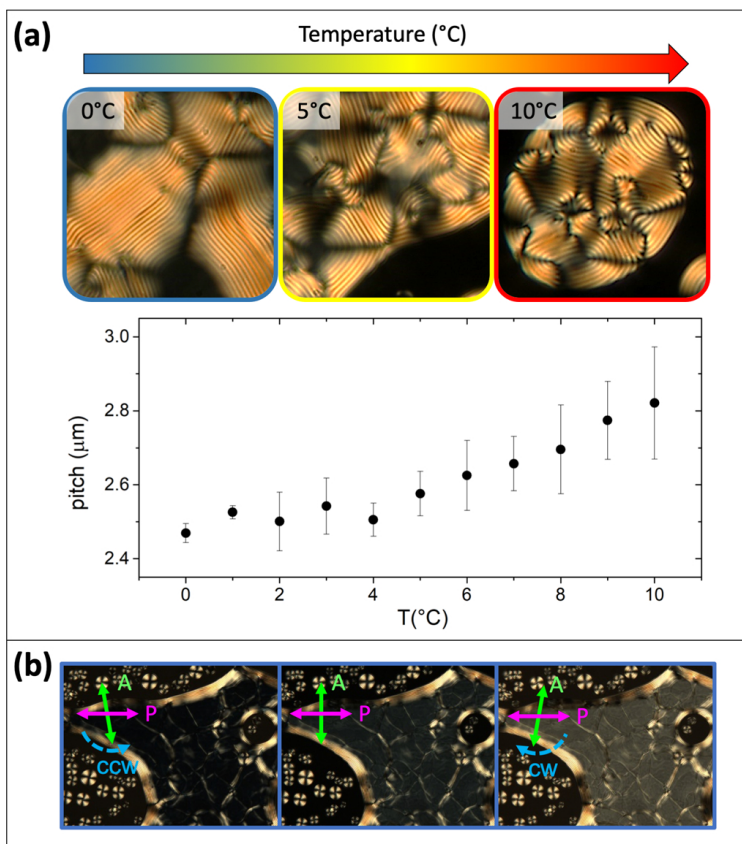
Fig. 2.20 3pDD cholesteric phase characterization at ≈ 500g/l. (a) The phase cholesteric pitch is in the infrared range of the light spectrum and grows with temperature, producing (b) a left-handed macrohelix.

It is important to stress that 3pDD liquid crystals are produced only at an extremely high concentration in order to overcome the destabilization in the interactions among their termini. Such a high concentration can be reached thanks to a very high solubility, presumably given by the greater extent of ionization of these molecules, which plays a key role in this story, even if not fully understood at this time.

A comparison of the complete phase diagrams of the three species is depicted below:
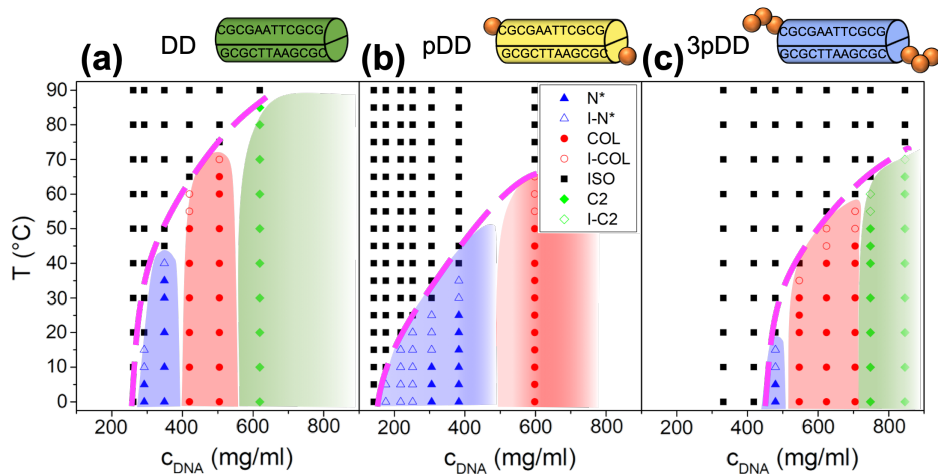
Fig. 2.21 Comparison of DD, pDD and 3pDD phase diagrams (T vs c_{DNA})

From this study it results that terminal modifications to nucleic acid oligomers alter the stability of their mesophases without totally disrupting them. In agreement with full atom predictions, we can see a trend of decreasing stability as more phosphate groups are added to esterify the 5'-hydroxyl group of the first nucleotide.

The addition of a triphosphate group opens up the possibility of generating ligation products without the necessity of adding a condensing agent such as EDC, as shown for a complementary templating sequence by the group of Jack Szostak.[30] Unfortunately, we could not detect ligation products in our samples. It is indeed important to take into consideration that we have been working with the sodium salt of oligonucleotides, and that the triphosphate ligation reaction is catalyzed by divalent metal ions such as magnesium and manganese. Future efforts must be put into synthesizing magnesium salts or into exchanging the counterions without disrupting the fragile triphosphate group. A successful attempt in this direction might open up the production of long covalently ligated polymers.

2.9 Removing the lower size limit for nucleic acids liquid crystal phases

Since the discovery of chromonic behavior in short mixtures of nucleic acids double helixes, a great effort has been put in finding the lower size limit for the transition to liquid crystal state, published first for DNA sequences with a length equal or larger than 6nt.[22] Since a main focus of this whole research project is to define whether a *liquid crystalline origin of Life* could be feasible, we must define the minimum size required for the supramolecular assembly and determine whether it is compatible with prebiotic synthesis.

The classic model describing chromonic liquid crystals of nucleic acids is based on the staking of double helices, that are not expected to form in solution for a length below 6nt, according to well-characterized nucleic acid thermodynamics.[11] Nevertheless, liquid crystals have been observed in tetramers of DNA having self-complementary sequences GCCG and ATTA.[54]

These molecules are not present as double helices in solution and to produce supramolecular order they must follow a different path. What it was shown is that these molecules organize through *running bonds* aggregation, meaning that they cooperatively pair and stack, switching directly from single stranded to supramolecular aggregates thanks to the extra gain of energy given by the coaxial stacking with the adjacent molecules.

Still, aside from G-quadruplex producing molecules,[55] no nucleic acid sequence shorter than four nucleotides has ever been shown to organize in a liquid crystal state. Molecules such as dinucleotides would require such a high concentration to undergo the same *running bonds* aggregation that they produce amorphous precipitates before reaching it and transitioning in a liquid crystal phase.

The characterization of a peculiar mechanism such as the *running bonds*

aggregation for liquid crystal formation, together with the finding of liquid crystal order at very high concentration in 3pDD pushed us in the direction of studying mixtures of triphosphorylated nucleosides (NTPs), also attracted by their intrinsic reactivity due to pyrophosphate behaving as a good leaving group in a phosphoryl transfer reaction.

What we found is that liquid crystals are produced in mixtures of ATP and TTP or GTP and CTP, but no mesophase is observed in pure ATP, TTP and CTP solution. GTP just by itself or mixed with other NTPs produces very stable liquid crystals assembled from G-quadruplexes. This set of observation is summarized in the following figure:
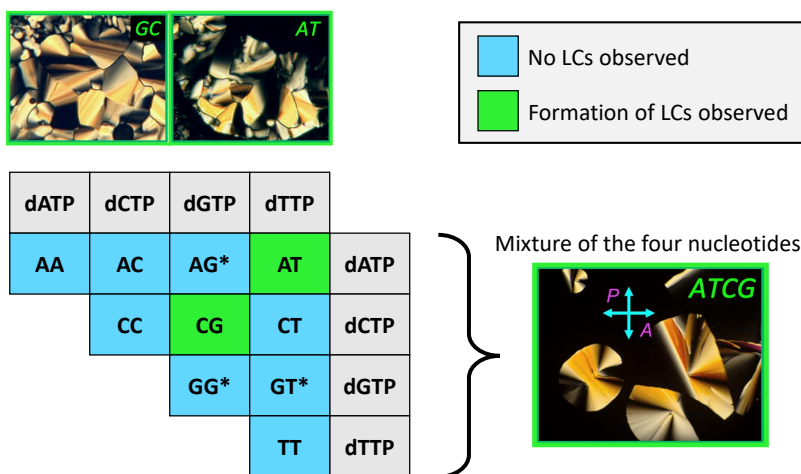
Fig. 2.22 Liquid crystals in solutions of NTPs.

These columnar phases appear at strikingly high concentrations if compared to longer sequences previously studied. A comparison of the phase diagrams of liquid crystals produced by oligomers of different sizes and terminal modifications is shown:
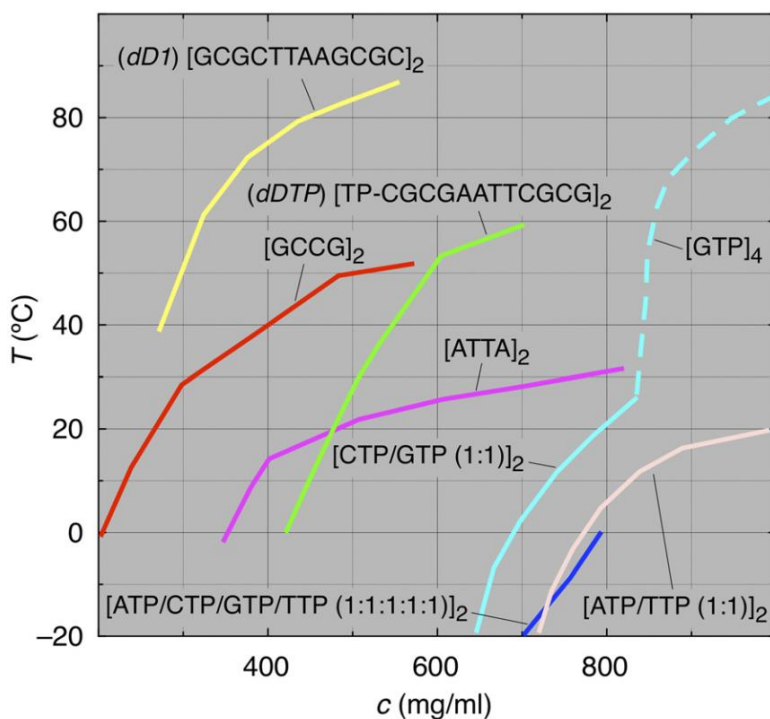
Fig. 2.23 Liquid crystals phases in nucleic acids of different lengths and terminal modifications.

To gain insight in the structure of these mesophases, X-ray scattering experiments have been performed by Gregory Patrick Smith to show that:

> **i:** NTPs are organized as pairs when in liquid crystal state, with the exception of GTP by itself or GTP:CTP mixtures at very high concentrations, where G-quadruplexes are responsible for the formation of supramolecular aggregates (in agreement to the phase diagram shown in Fig. 2.23).
>
> **ii:** the NTPspairsare stacked with the same spacing that it is measured in a conventional double helix of DNA.

These surprising results support the idea that even when they are not linked by any backbone, NTPs will pair following Watson-Crick selectivity among all the possible configurations shown in Fig. 1.3 to produce a double-helix-like structure kept together by weak bonds.

As for the case of DNA tetramers, thepairsdo not exist in solution because of their intrinsic instability. Their formation is the result of a cooperative phenomenon where production of pairs and their organization in linear aggregates happens simultaneously, transitioning from single triphosphorylated nucleosides to chromonic aggregates thanks to the gain in hydrogen bonding together with stacking energy.

Attempts at producing liquid crystals with nucleosides or nucleoside monophosphates at this day has given no positive results. These species exhibit a lower solubility due to the less extent of ionization, a subtle difference that could reasonably cause this qualitative difference. Nevertheless, it has to be noticed that among the commercially available NTPs that we could try, not everyone has been shown to produce liquid crystals. This observation underlines how critical for this small molecular species are the history of the samples or the presence of different counterions or pH with regard to liquid crystal ordering. It is now clear that future efforts will be put in standardizing the samples and comparing them in a more rigorous way.

2.10   2MeImpG liquid crystal phases and chemical reactivity

Given the limitations of EDC with respect to liquid crystals phases, and with the will to study a system with a more straightforward chemistry and a greater compatibility with prebiotic conditions, we decided to put aside 5'-phopshate nucleic acids and investigate to 2-methylimidazole-activated nucleotides, trying to understand whether these molecules can produce liquid crystals regardless of the altered charge distribution and the presence of the imidazole moiety.

2MeImpG is the activated form of guanosine monophoshate and it is known to react through a complex mechanism elucidated by Szostak et al.[31] and shown in Fig. 2.24
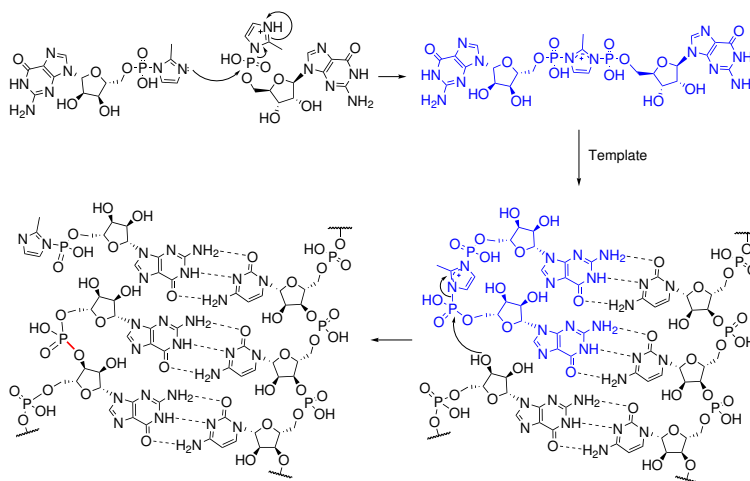
Fig. 2.24 Mechanism for primer estension by 2MeImpN.

The reaction proceeds first with the formation of a imidazolium-bridged dinucleotide starting from two 2MeImpN molecules with a 2-methylimidazole as a byproduct. This intermediate interacts though stacking and hydrogen bonding with a template and primer (see Fig. 2.24a). The primer 3'-hydroxyl group attacks the phosphate group of the proximal reactive nucleotide resulting in the elongation of the primer and in the formation a 2MeImpG as a leaving group, which can react further (see Fig. 2.24b). In-

terestingly, the imidazole-based activations show a strong regioselectivity towards 5'-3' linkages during primer extension experiments, making them a favorable choice in the context of the origin of Life.[56]

Among the four nucleotide derivatives we decided to start our investigation with 2MeImpG, since GMP is known to produce supramolecular structures by itself,[57] namely G-quadruplexes, and organize as a liquid crystal.

We began our study trying to understand through optical techniques whether 2MeImpG, which was synthesized by Federico Caimi as a disodium salt according to Joyce et al.[58] and Walton et al.[31] is capable to form supramolecular structures. Since Ethidium Bromide has a higher quantum yield when intercalated into G-quadruplex structures[59] we used a noncarinogenic derivative called GelRed to probe the presence of supramolecular structures, using rTMP and 2-methylimidazole as controls.
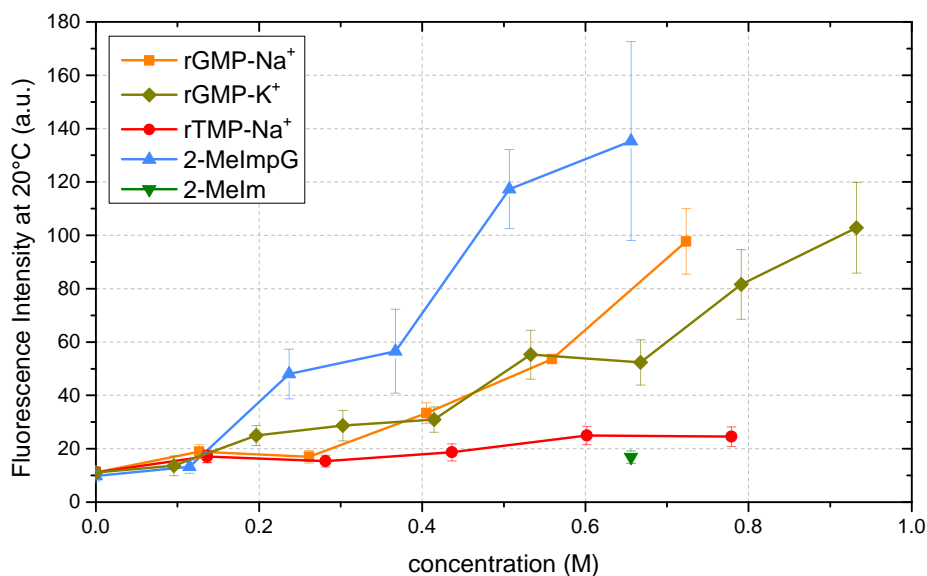
Fig. 2.25 Fluorescence emission of GelRed in rGMP, 2MeImpG and rTMP mixtures.

Results in Fig. 2.25 were promising and suggested that the presence of the 2-methylimidazole does not disrupt the capability of 2MeImpG to form G-quadruplexes. At very high concentrations rGMP-Na$^+$ and rGMP-K$^+$ produce birifringent crystalline-like structures, while 2MeImpG produces amorphous precipitates. It is well known that the solubility of the acidic form of nucleosides is very reduced, making them not suited for the formation of liquid crystals, where the molecules must stay in solution up to the concentration needed for the phase transition. Even when the nucleosides are ionized, the pairing with different counterions could have a huge impact on their behavior, we thus wanted to find the best choice for our purposes. Our results are tabulated as follows, where solubility is given as the concentration of nucleoside in the soluble phase after it reaches saturation concentration, as estimated through UV absorbance using rGMP exctinction coefficient:[60]

Table 2.2 Nucleic acid solubility

Compound	Counterion	Solubility
rGMP	Na$^+$	$\approx 250 g/l$
rGMP	K$^+$	$\approx 440 g/l$
rGMP	NH$_4{}^+$	$> 500 g/l$
2MeImpG	NH$_4{}^+$	$\approx 320 g/l$
2MeImpG	NH$_4{}^+$	$> 500 g/l$

Among the guanosine derivatives, only GMP (deoxyguanosine monophosphate) was reported in literature for producing nematic liquid crystals at above 70g/l and columnar at higher concentrations,[55] while we reported columnar liquid crystals formation in GTP-Na$^+$.

It is important to notice that at neutral pH rGMP has a net charge of -1.5, while GTP has a net charge of -3.5 and 2MeImpG a net charge of -0.5 due to the 2-methyilimidazole behaving as a weak base.

To our great suprise, we could find liquid crystalline nematic phases

also in 2MeImpG-NH$_4^+$, whose textures and behavior are remarkably close to the ones found in rGMP-NH$_4^+$.
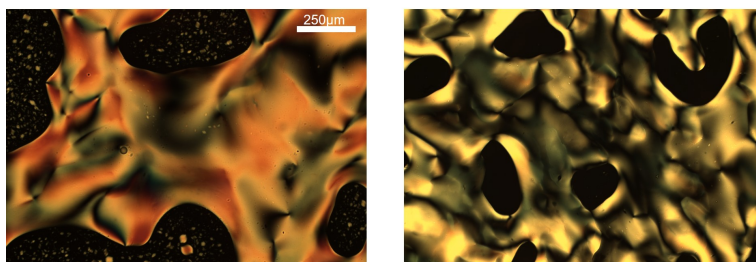
Fig. 2.26 2MeImpG-NH$_4^+$ nematic phase at 75g/l with sample thickness equal to $20 \mu m$ (left) and $5 \mu m$ (right).

Following this observation we produced a complete phase diagram, which resembles very closely the rGMP one, making it reasonable to assume that both are producing a nematic liquid crystal of G-quadruplex stacks as reported in literature for rGMP, while the reduction in stability of the 2MeImpG phases at high concentration could be due to an interference from the 2-methylimidazole group hanging out of the tetrads columns.
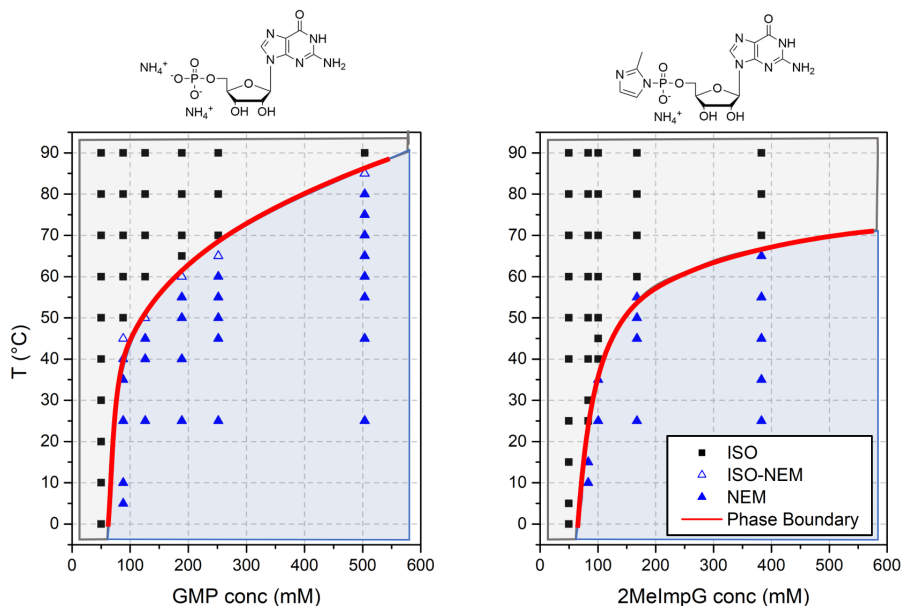
Fig. 2.27 rGMP-NH$_4^+$ and 2MeImpG-NH$_4^+$ phase diagrams comparison.

Given that the spacing between the adjacent base pairs of a G-quadruplex is the same as the one measured in a canonical DNA or RNA double helix, we expect the reaction described by Szostak group for a primer extension to occur also in our solutions.

We performed an NMR experiment on a Bruker 500 Avance III (11.74 T) acquiring the spectrum of the ^{31}P decoupled from proton. The samples were prepared at a concentration compatible with liquid crystals formation, equal to 150g/l. As previously shown, ammonium salts of nucleosides produce nematic liquid crystals (dark circle), while sodium salts of nucleosides remain in an isotropic state and are thus a very convenient negative control for the role of liquid crystals on the chemical reactivity of the system (light circle).
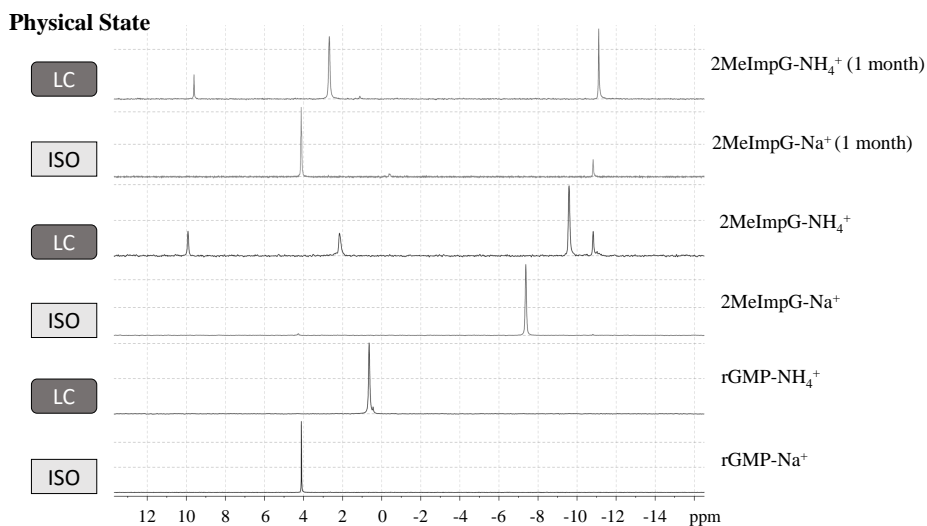
Fig. 2.28 ^{31}P-NMR spectra for 2MeImpG and rGMP samples.

NMR spectra for 2MeImpG shown in Fig.2.28 are obtained diluting the samples down to 5g/l either at time zero or after one month at room temperature. We can clearly detect the imidazolium-bridged dinucleotide at $\approx -11 ppm$. As expected the isotropic mixture (2MeImpG-Na$_+$) results mostly in the hydrolysis product rGMP-Na$^+$, with a small amount of

imidazolium-bridged dinucleotide still present. In the liquid crystalline state the reaction has a completely different products outcome and kinetic. At time zero we can already detect four peaks, assigned to the 2MeImpG-NH$_4^+$ at $\approx -10 ppm$, and the imidazolium-bridged dinucleotide at $\approx -12 ppm$. Two peaks at positive ppm could not be assigned but might correspond to polymeric states of the nucleoside. After one month of reactions, surprisingly the hydrolyzed rGMP is barely detectable, while the peak at $\approx 2 ppm$ becomes dominant together with the signal corresponding to the imidazolium-bridged dinucleotide, which seems to be surprisingly protected by hydrolysis.

To understand how it is possible that so much activated imidazolium-bridged dinucleotide is present in solution without either hydrolyzing nor getting consumed to push forward the polymerization reaction we decided to design a model for the reaction on a G-quadruplex structure. Since the reaction mechanism is quite complex and involves multiple steps it is not trivial to analytically predict whether the reaction can propagate to long products or whether it is destined to produce small polymers regardless of the presence of reactive molecules still in solution.

A simple and straightforward simulation (Appendix A) whose results can be seen in Fig.2.30 was set up in Matlab to understand the possible outcome of the chemical reaction of 2MeImpG, templated by the G-quadruplex supramolecular structure. The simulation is based on an array of size n representing the 2MeImpG molecules aggregated in a supramolecular structure of n elements. A second array b stores the information on chemical bonds (0 no bonds, 1 imidazole bridge, 2 phosphodiester bonds). The simulations are run for a number of temporal steps sufficient to reach equilibrium (no evolution for at least 90% of simulation time) and for each temporal step a molecule is selected and the system is allowed to evolve according to the model proposed by Szostak group. Four rules are es-

tablished to account for all the possible cases (Fig.2.29). Each reaction depicted can occur in both directions.) and give as an outcome a product distribution whose largest product is an hexamer. This product distribution shown here is independent by the choice of the n parameter as long as it is higher than ≈ 15. This simulation suggests that as long as the system is static and the 2MeImpG molecules cannot diffuse in and out of the aggregates, the product size is limited by the very mechanism of reaction (Fig.2.30) and independent from the size of the initial aggregates.
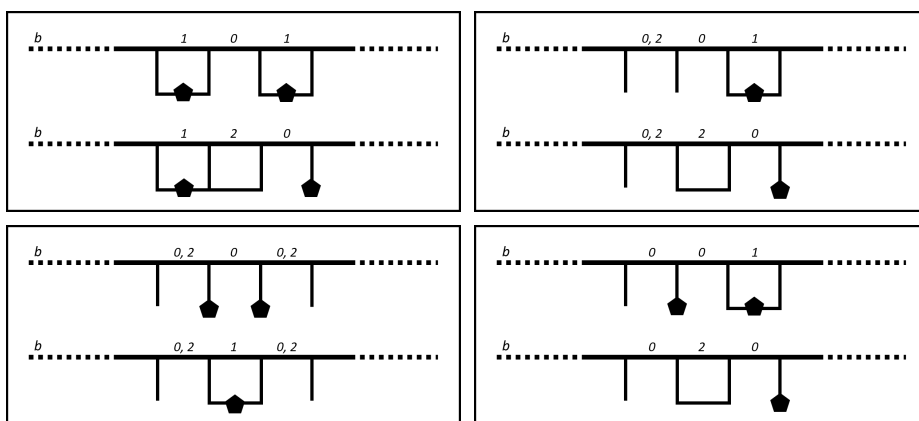
Fig. 2.29 Four cases taken into account for evolution of the 2MeImpG reaction simulation.
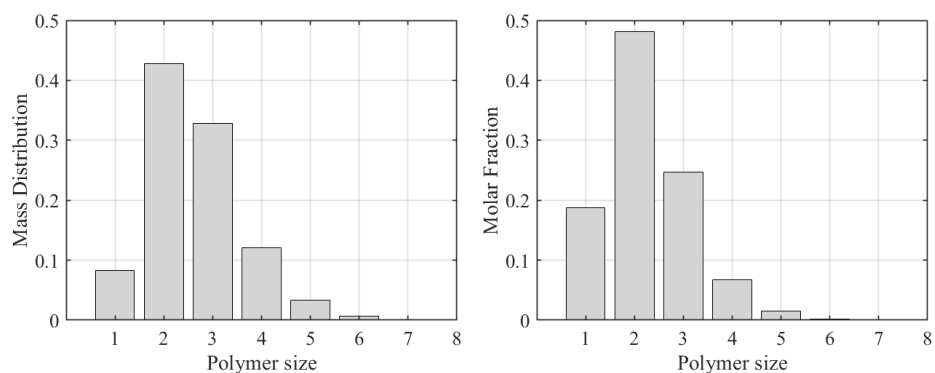
Fig. 2.30 Simulation for the 2MeImpG reaction occurring in a static G-quadruplex structure. Simulation was performed on aggregates of 500 elements for 100000 time steps and averaged on 100 simulations.

We ran a negative ESI mass spectrometry experiment to obtain information on the identity of the reaction products. In Fig.2.31 the spectra obtained this way are shown together with a sketch of polymeric state as a function of m/z.

Polymerization products up to pentamers can be detected in the ammonium salt of the nucleoside, that was reacting in a liquid crystalline state. Interestingly, this result is consistent with our simulation outcome, supporting the idea that diffusion is not permitted in this system, possibly due to very strong stacking interaction between the large hydrophobic G-quartets.

On the contrary, the sodium salt of the nucleoside prepared at the same concentration while being in an isotropic state could not polymerize significantly over trimers, highlighting the templating effect of liquid crystals. These results show how crucial it could be the dynamic in these supramolecular structures, whose potential could be optimized tuning their stability and allowing molecules exchange over time.
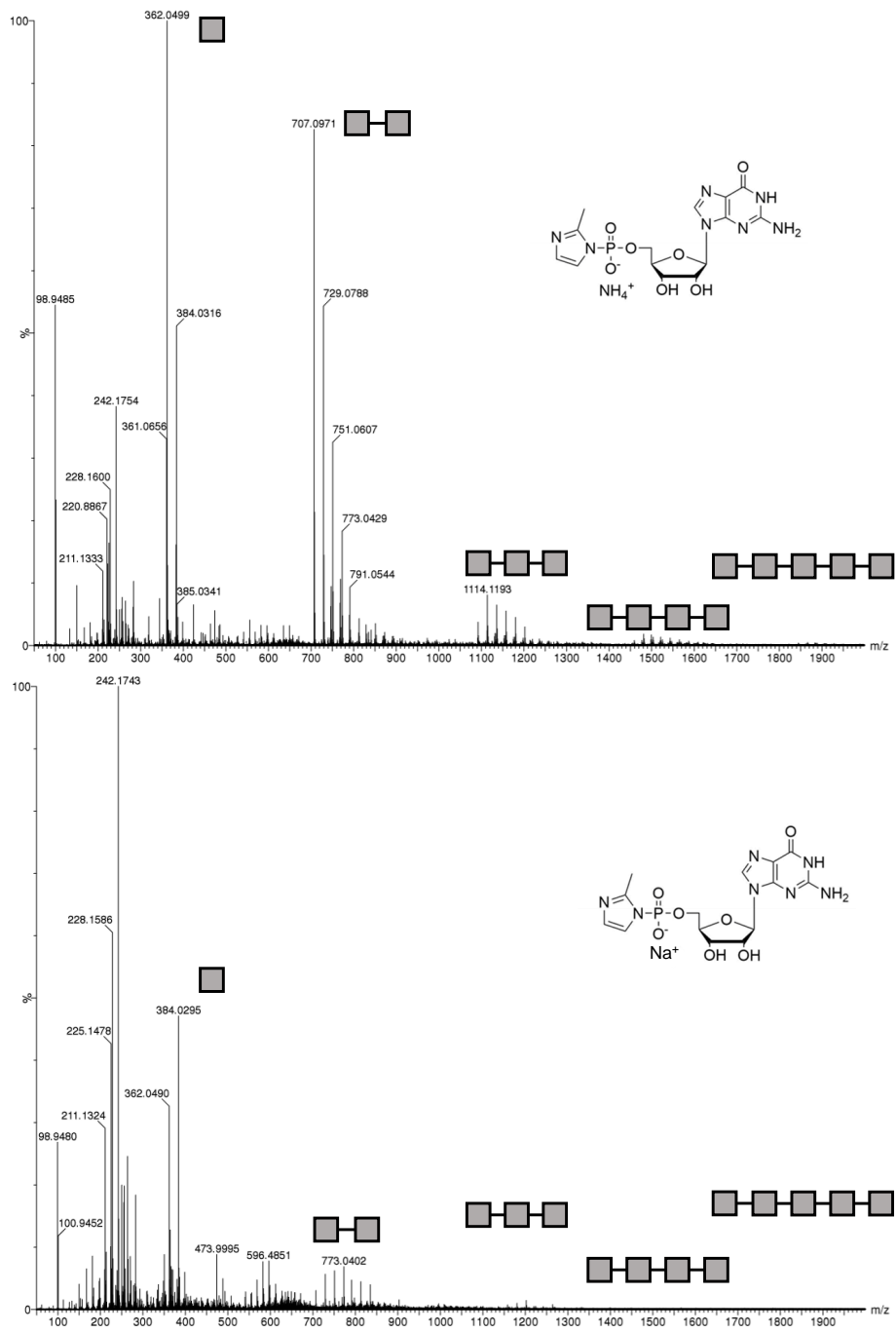
Fig. 2.31 ESI Mass Spectrometry of 2MeImpG over time.

2.11   Dynamics in the DNA liquid crystals

Being the liquid crystal phase of nucleic acids a dynamic system, we were interested in understanding the residence time of a DNA molecule in its aggregate. This number can be of critical importance for the outcome of chemical reactions such as the one described in Chapter 2.10, since to drive the formation of very long polymers, the reactive groups have to be regenerated, a phenomenon impaired in a static system.

To study the dynamics in our system we doped various solutions of DNA at different concentrations and temperatures using $200\mu M$ of a Fluorescein-tagged DNA sequences (purchased from IDT, HPLC purified). We started studying the DD sequence, blunt ended dodecamer interacting weakly only through hydrophobic effect and capable to produce liquid crystals:

C G C G A A T T C G C G | C G C G T T A A C G C G |
G C G C T T A A G C G C | G C G C T T A A G C G C |

All non fluorescent DNA samples used in this work have been dialyzed against salted water in order to obtain a fixed $\frac{[NaCl]}{[DNA]} = \frac{1mol}{300g}$.
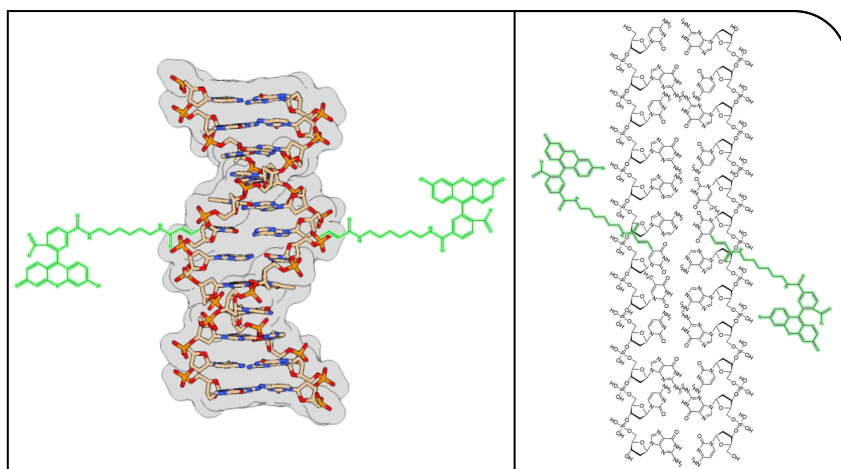
Fig. 2.32 Left panel, sketch of DD$_{\text{iFluorT}}$ double helix. Right panel, structural formula of DD$_{\text{iFluorT}}$ double helix.

To avoid impairment of the end-to-end interactions, at the basis of liquid crystals formation, the fluorescent moiety is added at the center of the sequence as a chemical modification to a thymine base (*iFluorT* modification from IDT catalog). We started our study characterizing the diffusion of DD$_{iFluorT}$ molecule performing a series of FRAP experiments. During a FRAP experiment, the fluorophores located in a small spot of the sample are bleached using a high energy source and a high magnification objective. The difference in concentration of fluorescent molecules created this way is then dissipated over time through a diffusive process that can be analyzed to gather information on the self-diffusion rates. The mathematical description of FRAP is closely related to the one of molecular diffusion in two dimensions starting from a configuration (at t=0) in which all molecules are concentrated in a single point (r=0). As the molecules diffuse we can define a probability distribution to find a molecules over time (t) and distance (r). The same probability distribution can be used to describe the bleached spot evolution in a FRAP experiment:

$$P(r,t) = e^{-\frac{r^2}{4D \cdot t}} \cdot \frac{A}{4\pi D \cdot t} \qquad (2.10)$$

Where the first exponential term stores a Gaussian function whose width is controlled by the diffusion coefficient D, while the second term is a rational function describing the Gaussian curve amplitude evolution. Unfortunately, while visually appealing and easy to interpret, for very low concentrations and high temperatures (i.e. low viscosity) the amplitude evolution for the spot over time has a poor signal to noise ratio and can be heavily influenced by drifts such as bleaching over acquisition. The best approach to extract diffusion coefficients in our system it is to extract the information from the evolution of the waist of the bleached spots over time. In order to do so, we can fit the radial average of our bleached spots with a Gaussian function in the form of:

$$A \cdot e^{\frac{-(x-x_0)^2}{2 \cdot r^2}} + y_0 \qquad (2.11)$$

We can see from Eq. 2.10 that the evolution of r^2 over time is linear and its slope can be fitted to easily obtain diffusion coefficients, given that $r^2(t) = 2 \cdot D$. An example of the waist data obtained fitting the spots profile with this approach can be seen below:
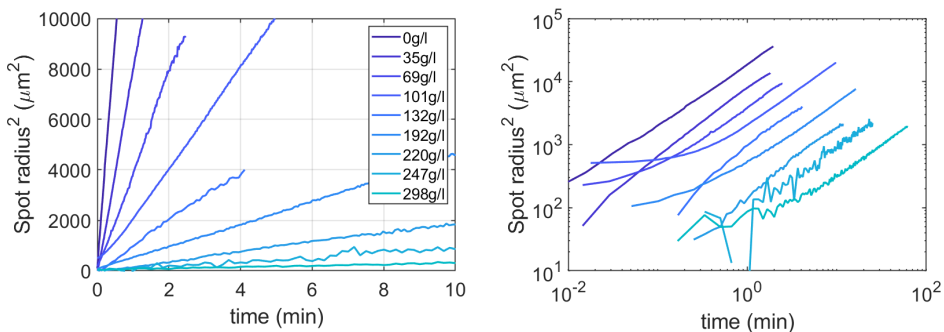
Fig. 2.33 Bleached spot waist evolution over time in DD at 5°C and different concentrations. Right panel shows the same data using a logarithmic scale.

As discussed in Chap. 2.8, DD does transition to a liquid crystalline phase above $\approx 225 g/l$. Theoretical models for nematic liquid crystal transition describe the phenomenon as rising from the growth of linear aggregates of double helices,[23] whose size slowly increases with concentration and abruptly grows upon phase transition. Accordingly, our data show a great reduction in diffusion rates of $DD_{iFluorT}$ as DD concentration increases, as shown in Fig. 2.33.

Since we are lacking a complete theoretical model to evaluate the nature of the slow-down measured in our system, we decided to study a series of samples prepared using the following DNA sequences:

sDD sequence, sticky ended dodecamer interacting more strongly than DD due to its terminal self-complementary overhangs and capable as well

to produce liquid crystals with similar stability as DD:

G C G A A T T C G C G C|G C G A A T T C G C G C|
C G **C G C T T A A G C G|C G C G C T T A A G C G|**

TT sequence, a 14nt long oligonucleotide producing 12bp long double helices with two non-complementary overhangs. It cannot produce liquid crystals and it is not expected to participate to an end-to-end aggregate:

C G C G A A T T C G C G T T
T T **G C G C T T A A G C G C**

FT sequence, a 12nt long oligonucleotide producing 8bp long double helices with four non-complementary overhangs. It cannot produce liquid crystals and it is not expected to participate to an end-to-end aggregate:

T T **C G C A T G C G** T T
T T **G C G T A C G C** T T

The diffusion coefficients for either $DD_{iFluorT}$, $TT_{iFluorT}$ and $FT_{iFluorT}$ in DD mixtures at varying concentration are shown below:
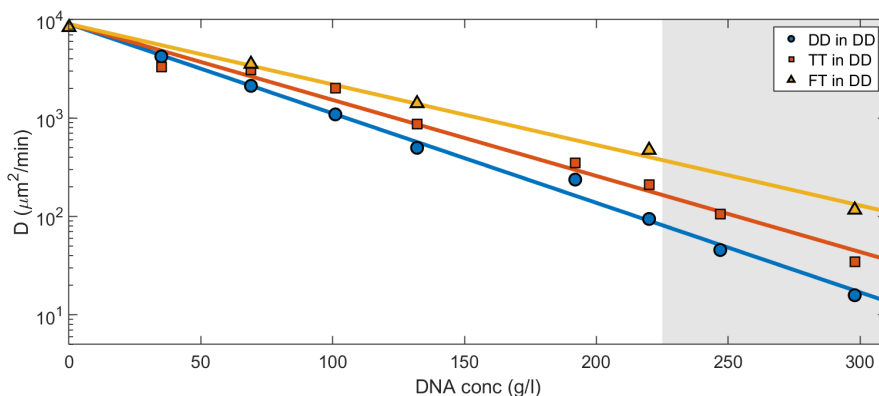
Fig. 2.34 Diffusion coefficients measured for different fluorophores added to a solution of DD at a concentration ranging from 0 to 298g/l. Shaded area represents liquid crystalline region.

Interestingly, the data obtained this way show an exponential behavior

and no clear sign of a transition when the system becomes nematic. By a direct comparison between the three data sets, we can see that $DD_{iFluorT}$ mobility decreases much faster than the one of both $FT_{iFluorT}$ and $TT_{iFluorT}$, as expected if terminal attractive interactions are present between the fluorescent probes and the matrix. Unexpectedly, $TT_{iFluorT}$ molecules, which were designed to be non-interacting with DD, diffuse more slowly than $FT_{iFluorT}$. A straightforward interpretation of these data is that both $DD_{iFluorT}$ and $TT_{iFluorT}$ have an additional attractive interaction with the surrounding matrix made of DD, which causes a slow-down of their diffusion rate. The great slow-down that we measure for $DD_{iFluorT}$ confirms that (i) the fluorescent moiety is not impairing the stacking interaction and (ii) FRAP can be exploited to characterize end-to-end interactions among short nucleic acid molecules. Moreover, according to these results, a certain degree of terminal interaction is present between $TT_{iFluorT}$ and DD.

By fitting these data altogether we can have a more precise estimate of the diffusion coefficient of the different fluorophores at the low concentration limit. This value allows us to extract the hydrodynamic radius (R_h) of the molecules from Stokes-Einstein equation:

$$D = \frac{RT}{N_A \cdot 6 \cdot \pi \cdot \eta_w \cdot R_h} \tag{2.12}$$

Where η_w is water viscosity at 5°C. The hydrodynamic radius calculated this way is equal to 1.16nm, a precious parameter for future modeling.

Molecular diffusion can be described as an activated process, where

$$\frac{D}{D_0} = A \cdot e^{\frac{Ea}{RT}} \tag{2.13}$$

from which it can be derived:

$$ln\left(\frac{D}{D_0}\right) = ln(A) - \frac{\Delta S}{R} + \frac{\Delta H}{RT} \tag{2.14}$$

Where D_0 is the diffusion coefficient in the diluted regime. An Arrhenius plot of the diffusion coefficients measured at a fixed concentration and varying the temperature should thus give us access to the enthalpies of activation:
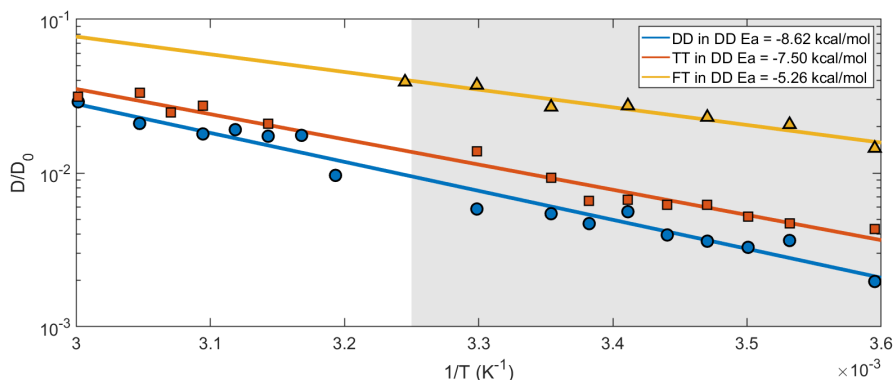
Fig. 2.35 Arrhenius plot for the diffusion coefficient of different fluorophores added to a solution of DD at 298g/l varying the temperature. Shaded area represents liquid crystalline region. Enthalpies of activation extracted from linear fits are shown in legend.

As expected from Eq.2.14, the data show a linear behavior and the points can be linearly fit to obtain values for enthalpies of activation. In agreement with the previous measurements as a function of concentration, we determine $\Delta H^{\ddagger}_{FT} < \Delta H^{\ddagger}_{TT} < \Delta H^{\ddagger}_{DD}$. Once again, we see no clear transition in the molecules mobility when the matrix becomes liquid crystalline.

To further corroborate FRAP as a tool to measure weak attractive interactions between DNA double helices terminal, we measured diffusion rates for sDD, a molecule that produces liquid crystalline phases in the same regime as DD but it is expected to have a stronger end-to-end interaction due to its sticky ends.
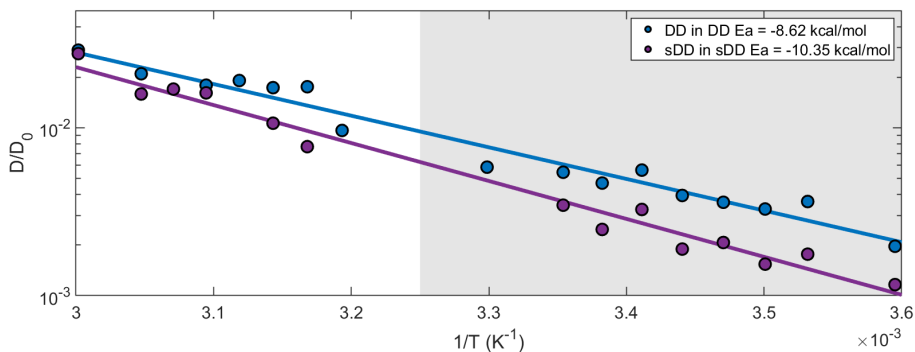
Fig. 2.36 Diffusion coefficients measured for $DD_{iFluorT}$ added to a solution of DD 298g/l and $sDD_{iFluorT}$ added to a solution of sDD 273g/l. Shaded area represents liquid crystalline region.

The characterization of $sDD_{iFluorT}$ in sDD shows that $\Delta H^{\ddagger}_{DD} < \Delta H^{\ddagger}_{sDD}$, as expected if correctly probing end-to-end attractive interactions.

This work places the basis for future developments in order to obtain more information on the dynamics in concentrated nucleic acids systems and their liquid crystals. Moreover, given the extreme complexity in manufacturing monodispersed nanometric particles with a charge density as high as DNA makes this work particularly interesting to understand fundamental interactions on a length scale otherwise very difficult to tackle.

To completely capture the contribution of the activation energy in this diffusive process, a set of data as a function of both temperature and concentration has to collected for the various molecules used in this work. A preliminary example of these data can be seen in the following figure:
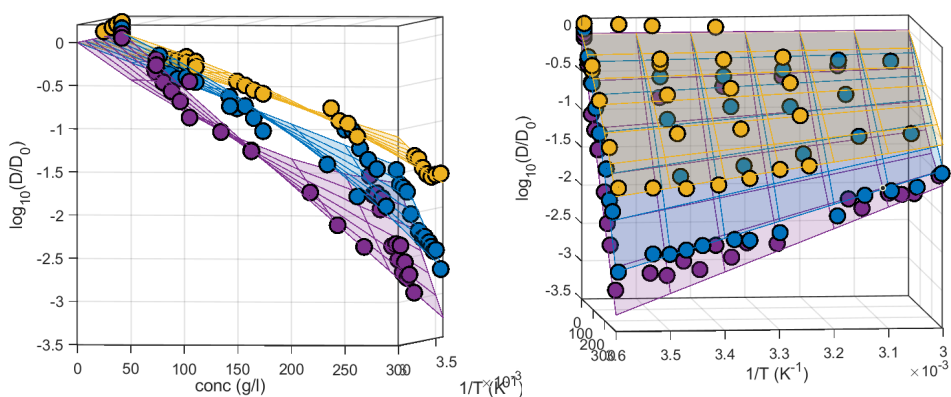
Fig. 2.37 Diffusion coefficients measured for $FT_{iFluorT}$ added to a solution of DD or sDD (yellow), $DD_{iFluorT}$ added to a solution of DD (blue) and $sDD_{iFluorT}$ added to a solution of sDD (violet). Surfaces are used to guide the eye.

It is important to underlie that we demonstrated how FRAP is a good technique to collect information on nucleic acids mobility and their end-to-end attractive interactions. Given the great reduction in mobility measured when stacking (using DD) and sticky (using sDD) interactions are introduced in the system, we should expect a great slow-down of dynamics in highly concentrated rGMP solutions producing G-quadruplexes, whose interactions are so strong that the solution undergoes nematic transition at a concentration as low as $\approx 35 g/l$ starting with molecules much smaller than the dodecamers used in this work. A direct result of this phenomenon could be the production of a static system affecting the outcome of 2MeImpG chemical reactions as described extensively in Chap. 2.10.

Future work should be aimed at (i) model the data to obtain a clear understanding of diffusive processes in systems composed by small nucleic acid molecules, (ii) understanding why from the point of view of the molecules mobility no discontinuity is observed when undergoing liquid crystalline transition and (iii) clarify how terminal interactions affect the dynamics in a liquid crystal phase, where chemical reactions occur more efficiently than in a isotropic solution.

CHAPTER 3

Conclusions and Future Perspectives

The results shown in this thesis strengthen the potential role of liquid crystal ordering in the emergence of Life on the prebiotic Earth. In this work it is shown how the supramolecular organization in nucleic acids is not just an exotic phenomenon that occurs under specific conditions but it is indeed rooted deeply in their physicochemical properties and resilient to chemical modifications or small perturbations: we could find liquid crystals in hydroxyl-terminated, monophosphate-terminated and triphosphate-terminate short oligonucleotides, triphosphorylated nucleosides and 2-methylimidazole-activated nucleosides. Still, up to this day no liquid crystal ordering in simple nucleosides mixtures has been reported, possibly due to their reduced solubility. Future efforts should be put in understanding how the features of these molecules affect the liquid crystal formation in order to obtain predictive tools in a multi-parameter space. To perform this, many variables have to be carefully checked such as pH and counterions nature and concentration, given their demonstrated impact especially on small molecules such as 2MeImpG and GMP.

We could show that in liquid crystals made of short RNA oligonucleotides, a non-enzymatic ligation chemical reaction is enhanced by the

physical proximity of double helices termini being part of the supramolecular assemblies stabilized by the supramolecular ordering. We could measure both a boost in reactivity and a tendency of the system to favor intermolecular reactions over intramolecular reactions, a reasonable outcome of the system symmetry breaking, which prevents polymers cyclization and favors the formation of long linear polymers, necessary to exhibit catalytic activity, carry genetic information and evolve towards Life.

Even though these results are the outcome of a study that relies on a condensing agent that is not prebiotically plausible, we must stress that they should be accepted as a proof of concept for the role of supramolecular ordering in the context of the origin of Life. Future studies could expand the present one by reproducing the data using cyanogen bromide or N-cyanoimidazole instead of EDC, to increase the relevance of our model from a prebiotic perspective.

Remarkably, we could detect polymerization products in a 2MeImpG reaction occurring in nematic liquid crystalline state. Analysis of reaction products through mass spectrometry reveals that indeed supramolecular ordering favors the formation of longer polymers with respect to an isotropic reaction, even though the reaction does not go to completion and the size of the products is modest. Our modeling of the system reveals that this outcome is expected for a static system, where diffusion of molecules inside and outside of the aggregates is not allowed. Future studies should expand the model to understand the relation between lifetime of molecules in the aggregates and reaction kinetics, and the same reaction should be studied at a higher temperature to tune these parameters or in a Watson-Crick bonded system to reduce the strength of interaction.

In light of these results it becomes tempting to justify the difference in the discontinuities measured for the EDC-mediated reaction yields (see Fig. 2.5) as a function of temperature when transitioning to a liquid crystalline state in mixtures of short RNA oligonucleotides as due to a different dy-

namic of the molecules, a fascinating hypothesis that should be explored in the future.

Finally, a novel approach based on FRAP measurements is described here to study the dynamics in solutions of short DNA oligonucleotides. This study produced a wide set of data that will be modeled to gain new insight in the properties of the system. A more complete characterization will be especially important in order to decompose the effects of physical bonds dynamics and transient diffusion between aggregates; a key element in determining the propagation of polymerization reactions to produce significantly long products.

Aside from these intriguing properties of nucleic acids liquid crystals, we must confront a major limitation in our system; the extremely high concentration required to trigger the phase transition. While it is true that nucleic acids liquid crystals emerge just by evaporation of diluted solutions, a careful analysis of the composition of prebiotic soups has to be done to determine whether the increase of concentration of other molecules in the surrounding could impair the formation of liquid crystals.

In the future more efforts have and will be put in expanding our knowledge of a potential liquid crystal origin of Life.

Materials and Methods

4.1 Microscope cells preparation

To prepare samples for microscope observations, so-called microscope cells, one or multiple drops of nucleic acid solution are pipetted on the clean surface of a glass slide up to a total volume of $0.5 - 1\mu m$. If a concentration higher than the stock solution is required, multiple droplets are pipetted on a glass slide, waiting for them to dry before any new addition. This dry material is finally resuspended in the appropriate volume of solution to obtain the desired target concentration directly on the slide. In order to observe these samples through standard bright field, polarized light and fluorescence microscopy, a second glass slide has to be placed on the top of the solution and is kept at a fixed distance of $20\mu m$ using silica spacer rods. In the samples prepared in this way, the LC domains extend for tens of micrometers and are thus generally much larger than both molecular and aggregate sizes and fill the space between the two glasses. The cells are sealed with Fomblin y25 fluorinated oil to mantain the sample concentration stable over long periods of time and prevent evaporation even at high temperatures. Samples prepared this way usually exhibit concentration gradient. Prior to every measurement and sample characterization, the microscope cells are thermally cycled to speed up diffusion and homogenization.

4.2 Microscope observations and phase diagrams measurement

Microscope cells were observed through a Nikon TE200 optical microscope through crossed polarizers to indentify and characterize birefringent textures. The microscope stage temperature is controlled using an Instec MK2000 peltier module. For phase diagram characterization the samples have been studied starting from low temperatures to avoid hysteresis and allowing them to equilibrate for 5 minutes or until the observed images were stable before annotations were taken on the system state.

4.3 1D PAGE

To resolve and quantify the reaction products of pRNA with EDC, $2\mu g$ of reaction product were dissolved in $18\mu l$ of total volume with $5\%\frac{v}{v}$ glycerol final concentration and $10\mu l$ of formamide. These samples were loaded in a 7 M urea denaturing 15% polyacrylamide gel (size $20x20x0.1cm^3$). Polyacrilamide gels were run in TBE buffer 1x in a Protean apparatus (Biorad) at 250V for approximately 2.5 hours. Staining was performed shaking the gels in a SYBR Gold 1x solution in TBE buffer for 5 min. Images were acquired in .tif format using a Gel Doc EZ Imager (Biorad).

4.4 2D PAGE

To perform a 2D PAGE it is necessary to run first a 1D PAGE as described above. When the run was completed, a single lane was excised from the gel using a scalpel. The lane was then rotated by 90°and placed on top of a second 7 M urea denaturing 20% polyacrylamide gel. The two gels were sealed together with 20% polyacrylamide gel and the second dimension was run for 4 hours at 300V. Staining and image acquisition were performed as described above.

4.5 Extraction of reaction yields from PAGE

PAGE images acquired through Gel Doz EZ Imager (Biorad) were processed in MacBiophotonics ImageJ software to extract lanes profiles as already described by Fraccia et al.[40] Since SyberGold was used to stain the gel, we can assume that the fluorescent signal imaged is linear with the moles of nucleic acids present in each band. This is no longer true for very short molecules, where the signal per gram of material drops dramatically, presumably due to a low efficiency of SyberGold intercalation. This phenomenon can be seen for example in 6SE, making the analysis of its reaction yield unreliable with this approach.

When the lanes are extracted as fluorescence intensity vs pixel position, the data are imported in MatLab and processed through a custom code to obtain quantitative information on reaction yields, defined as the fraction of phosphate termini reacted. Our algorithm proceeds as follow:

i) Using the DNA ladder as a reference we select the peaks corresponding to easily attributable pRNA sizes, usually 12nt, 24nt, 36nt and 48nt long polymers.

ii) Since the electrophoretic mobility of nucleic acids scales linearly with the logarithm of size of the molecule, we can obtain a relationship between the pixel position and the number of nt with the corresponding electrophoretic mobility.

iii) The lane profile as a function of nucleic acid size (nt) is integrated step-wise to obtain the amount of material at each polymerization step. **iv)** The normalized cumulative mass distribution though the lane is fitted with Flory Model for simple polymerization.[61]

4.6 Flory Model for simple polymerization

According to Flory Model[61] for a simple polymerization we can defined the total number of reactive monomers (in this case 12BE) present in solution as N_0 and N as the number of unreacted monomers when the reaction is

stopped. We can defined a reaction yield $p = \frac{N_0 - N}{N_0}$, which is equal to the probability of a phosphate terminus to react.

The probability $\pi(n)$ to produce a n-length polymer during the reaction can be defined as follows:

$$\pi(n) = p^{n-1}(1 - p)^2 \qquad (4.1)$$

Since each n-length polymer can be produced in n different configurations of the starting monomers, it follows that the probability to produce any -length polymer is:

$$P(n) = n \cdot p^{n-1}(1 - p)^2 \qquad (4.2)$$

Where P(n) is the number of monomers part of any n-length polymer over the total amount of monomers N_0, also defined as weight fraction distribution.

We can also defined a cumulative weight fraction distribution over polymers up to length i, that can be fitted to our data:

$$C(n) = \sum_{n=0}^{i} 1 - p^n - np^n + np^{1+n} \qquad (4.3)$$

4.7   pH determination

Since EDC reactivity is heavily dependent on pH, in order to characterize the kinetic of EDC condensation it has been necessary to measure the pH of our working solution. Given the very small volumes used, limited by the amount of material available, the pH of the mixtures containing 12BE together with PEG8k, $MgCl_2$ and EDC could be measured only using an Orion 9810BN Micro pH Electrode in a $3\mu l$ volume. The pH determined in this way is equal to 7.3 and the spectrophotometric measurements detailed in this study have thus been performed using an excess of HEPES pH 7.3

to match the one of the more complex mixtures.

4.8   FRAP-based measurement of viscosity

All FRAP experiments have been performed on microscope cells using a Nikon TE200 optical microscope equipped with a Fluorescein filter cube from Nikon. A LED source (CoolLED pE-4000) at 470nm was focused over a sample area with radius ranging from $50\mu m$ to $400\mu m$ using a high magnification objective (20x or 50x) to bleach the fluorophore. Recovery of fluorescence was recorded using a Nikon DS-Fi3 camera through a low magnification objective (4x or 10x) and using neutral filters to reduce bleaching during the acquisition step.

The time series obtained this way were processed subtracting a pre-bleached image and finding the bleached spot center at the first frame by fitting the image intensity with a Gaussian function. Once the position of the center of the bleached spot is defined, at every time step the subtracted images are radially averaged and a Gaussian function is fitted to extract waist size and amplitude at each frame.

For the FRAP experiments described in Chap. 2.4 we used Sodium Fluorescein as a probe to measure the viscosity of RNA and RNA-PEG mixtures due to its tendency to photobleach and due to its hydrodynamic radius (r = 0.45 nm) being comparable to the one of EDC. For 12BE and PEG mixtures, diffusion coefficients were calculated according to Soumpasis model[62] as $D = 0.224\frac{r^2}{t_{\frac{1}{2}}}$, where $t_{\frac{1}{2}}$ is the time after recovery begins at which the fluorescent signal measured in the center of the bleached spot is equal to half of the intensity in the same spot before the bleaching process.

For the FRAP experiments described in Chap. 2.11, the evolution bleached spot intensity was radially averaged and fitted at every frame using a Gaussian model as $A \cdot e^{\frac{-(x-x_0)^2}{2 \cdot r^2}} + y_0$ to extract the parameter r. The evolution or the spot radius over time was fitted to obtain diffusion coefficient given that $r^2 = 2 \cdot D \cdot t$.

4.9　Enzymatic Digestion of 6SE

6SE ligation products with EDC were enzymatically treated with SVP, Snake Venom Phosphodiesterase I extracted from *Crotalus adamanteus* (Sigma Aldrich). This enzyme has both 3'-exonucleolytic and endonucle- olytic activities but it is reported to be less effective on circular oligomers than linear oligomers.[51,52] A stock solution of SVP was produced resus- pending the lyophilized enzyme in 4 ml of reaction buffer (Tris-HCl 10mM pH 8.8, NaCl 25mM, MgCl$_2$ 10mM). To treat 6SE samples, $1\mu l$ of stock SVP was added to a $9\mu l$ solution of the reaction products diluted in reaction buffer to reach a concentration $c_{RNA} \approx 0.1 - 0.2 g/l$ and an enzyme activ- ity corresponding to $\approx 0.01 \frac{U}{ml}$. The enzymatic reaction was run at room temperature and quenched adding $2\mu l$ of 55mM EDTA pH 8 solution after 5', 10', 30' and 60' to the mixtures and immediately storing them at -20°C.

4.10　Salt exchange

Commercially available nucleosides and oligomers are generally provided as highly soluble sodium salts. In order to replace sodium ion with another ion of choice it is necessary to pack a resin column as follows: For 50mg of compound it is necessary to weight 7.5g of Dowex 50WX8 resin. This is a strong acid resin that has to be equilibrated against a strong base conjugated with the ion of choice. We tested this method using KOH and NH$_4$OH in order to substitute the sodium counterions with potassium and ammonium.

50mg of nucleic acid bear roughly 0.15mmol of counterions that have to be replaced. In order to do so, we have to use an excess of resin func- tional groups. The resin has a capacity of 1.7 meq/ml, and since 7.5g of Dowex 50WX8 resin can be packed in roughly 6ml, it will be capable to exchange 10mmol of ions, two order of magnitues more than necessary.

To equilibrate the resin acid functional group and make a salt of the

desired counterion we stir it gently in 75ml of a 2M strong base solution. This solution contains 150mmol of the desired ion, tenfold more than the acidic group on the resin.

We let the resin stir for at least 3 hours at room temperature and then wash it over a ROBU Filter-Funnel (# 21123) 125ml, porosity $16 - 40\mu m$. During this step we measure the pH of the flow-through, usually $\approx 700ml$ of milliQ water is enough to bring down the pH of the solution to neutrality, meaning that every acidic group is balanced and no excess of hydroxide should be present in solution.

When the resin is equilibrated at $pH \approx 7.0$, a 5ml syringe can be packed after the addition of a small amount of cotton at the bottom to prevent leakage. The resin is then rehydrated with $\approx$ 2ml of milliQ water and our nucleic acid solution is added at the top of the column as a 10g/l solution to begin the salt exchange.

We use the syringe plunger to speed up the process, collect the eluted solution and load it again for at least ten times. During this process, by the law of mass action the nucleic acid will bind the desired counterion. The collected solution can be lyophilized in a tared eppendorf to keep track of the amount of material recovered and its purity.

4.11 Solubility Measurement

To run solubility checks of rGMP and 2MeImpG we resuspended the lyophilized material at a high concentration ($\approx 500g/l$) in order to saturate the solution and produce a precipitate that is let equilibrate over time. The absorbance of the supernatant is measured and converted to a concentration using 12.08mM⁻¹cm⁻¹ as the exctintion coefficient reported by Cavaluzzi et al.[60] This value is used as a measurement of nucleotide solubility.

4.12 DD, sDD and FT samples preparation

DD and sDD oligonucleotides used in Chap. 2.11 were synthesized by Dr. Gregory Patrick Smith on an Äkta Oligopilot by Amersham Pharmacia Biotech following standard solid support approach. The oligonucleotides were resupended at a theoretical concentration of 100g/l and diluted down to 0.05g/l for UV absorbance quantification in a standard quartz cuvette of a 1cm path length. Exctintion coefficient were calculated according to Owczarzy[63] and samples concentrations resulted to be 73.8g/l for DD ($\epsilon_{260nm} = 110700$ M^{-1}cm^{-1}) and 77.4g/l for sDD ($\epsilon_{260nm} = 110300$ M^{-1}cm^{-1}), resulting in a mass excess that could be due to residual NaCl after salt precipitation. To get rid of the excess salt and make the two samples comparable, we ran a dyalisis purification using Slide-A-Lyzer kit from Thermo Scientific with 2k molecular weight cut off. 200 of each oligonucleotide were loaded diluted down to 50mM nucleotide concentration and dialysis was performed against 2L of 50mM NaCl milliQ water solution. Dyalisis solution was changed two times over two days and samples were liophylized and then resuspended to a final theoretical concentration of 50g/l. Again, UV absorbance measurements were performed to have a consistency check, resulting in a DD concentration of 36g/l and a sDD concentration of 33g/l, with a mass defect compatible with the salt added through dialysis.

REFERENCES

[1] M. Todisco, T. Fraccia, G. P. Smith, A. Corno, L. Bethge, S. Klussmann, E. M. Paraboschi, R. Asselta, D. Colombo, G. Zanchetta, N. A. Clark, and T. Bellini, "Nonenzymatic polymerization into long linear RNA templated by liquid crystal self-assembly," *ACS Nano*, vol. 12, no. 10, pp. 9750–9762, 2018.

[2] A. Lazcano, "Prebiotic evolution and self-assembly of nucleic acids," *ACS Nano*, vol. 12, no. 10, pp. 9643–9647, 2018.

[3] G. P. Smith, T. P. Fraccia, M. Todisco, G. Zanchetta, C. Zhu, E. Hayden, T. Bellini, and N. A. Clark, "Backbone-free duplex-stacked monomer nucleic acids exhibiting watsoncrick selectivity," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 115, no. 33, pp. E7658–E7664, 2018.

[4] R. Podgornik, "Sticking and stacking: Persistent ordering of fragmented DNA analogs," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 115, no. 35, pp. 8652–8654, 2018.

[5] S. Di Leo, M. Todisco, T. Bellini, and T. P. Fraccia, "Phase separations, liquid crystal ordering and molecular partitioning in mixtures of PEG and DNA oligomers," *Liquid Crystals*, vol. 45, no. 13–15, pp. 2306–2318, 2018.

[6] M. Todisco, G. Smith, and T. Fraccia, "Liquid crystal ordering of DNA dickerson dodecamer duplexes with different 5- phosphate terminations," *Mol Cryst Liq Cryst*, vol. 683, no. 1, pp. 69–80, 2019.

[7] X. Xia, *Bifacial Peptide Nucleic Acid (bPNA) as a Regulator of Nucleic Acid Function*. PhD thesis, 2015.

[8] Z. Shabarova and A. Bogdanov, *Advanced Organic Chemistry of Nucleic Acids*. 1994.

[9] N. B. Leontis, J. Stombaugh, and E. Westhof, "The non-watson-crick base pairs and their associated isostericity matrices," *Nucleic Acids Res.*, vol. 30, no. 16, pp. 3497–3531, 2002.

[10] 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*, vol. 23, no. 2, pp. 564–574, 2006.

[11] J. J. SantaLucia and D. Hicks, "The thermodynamics of dna structural motifs," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 33, pp. 415–440, 2004.

[12] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns," *Nature*, vol. 440, pp. 297–302, 2006.

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

[14] M. D'Abramo, C. L. Castellazzi, M. Orozco, and A. Amadei, "On the nature of DNA hyperchromic effect," *J. Am. Chem. Soc.*, vol. 117, pp. 8697–8704, 2013.

[15] I. J. Tinoco, "Hypochromism in polynucleotides," *J. Am. Chem. Soc.*, vol. 82, no. 18, pp. 4785–4790, 1960.

[16] H. Chen, S. P. Mesiburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollac, "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 3, pp. 799–804, 2012.

[17] G. Zanchetta, M. Nakata, M. Buscaglia, and N. A. Bellini, Tommaso Clark, "Phase separation and liquid crystallization of complementary sequences in mixtures of nanoDNA oligomers," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, no. 4, pp. 1111–1117, 2008.

[18] S. Asakura and F. Oosawa, "On interaction between two bodies immersed in a solution of macromolecules," *J. Chem. Phys.*, vol. 22, p. 1255, 1954.

[19] T. Soderberg, *Organic Chemistry with a Biological Emphasis, Volume II*. 2019.

[20] R. R. Rill, "Liquid crystalline phases in concentrated aqueous solutions of Na$^+$ DNA," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 83, pp. 342–346, 1986.

[21] L. Onsager, "The effects of shape on the interaction of colloidal particles," *Ann. NY Acad. Sci.*, vol. 51, p. 627, 1949.

[22] M. Nakata, G. Zanchetta, B. D. Chapman, C. D. Jones, J. O. Cross, R. Pindak, T. Bellini, and N. A. Clark, "End-to-end stacking and liquid crystal condensation of 6 to 20 base pair DNA duplexes," *Science*, vol. 318, no. 5854, pp. 1276–1279, 2007.

[23] C. De Michele, T. Bellini, and F. Sciortino, "Self-assembly of bifunctional patchy particles with anisotropic shape into polymers chains: Theory, simulations, and experiments," *Macromolecules*, vol. 45, no. 2, pp. 1090–1106, 2012.

[24] J. Lydon, "Chromonic review," *Journal of Mat. Chem.*, vol. 33, pp. 415–440, 2010.

[25] S. W. Fox, "Aleksandr Oparin," *Encyclopedia Britannica*, 2019.

[26] W. M. Powner, B. Garland, and J. D. Sutherland, "Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions," *Nature*, vol. 459, pp. 239–242, 2009.

[27] Z. Martins, O. Botta, M. L. Fogel, M. A. Sephton, D. P. Glavin, J. S. Watson, J. P. Dworkin, A. W. Schwartz, and P. Ehrenfreund, "Extraterrestrial nucleobases in the murchison meteorite," *Earth and Planetary Science Letters*, vol. 270, no. 1–2, pp. 130–136, 2008.

[28] M. H. Engel and N. Bartholomew, "Distribution and enantiomeric composition of amino acids in the murchison meteorite," *Nature*, vol. 296, pp. 837–840, 1982.

[29] J. P. Ferris, A. R. Hill, R. Liu, and L. E. Orgel, "Synthesis of long prebiotic oligomers on mineral surfaces.," *Nature*, vol. 381, pp. 59–61, 1996.

[30] R. Rajat, D. P. Bartel, and J. W. Szostak, "Kinetic and mechanistic analysis of nonenzymatic,template-directed oligoribonucleotide ligation," *Journal of the American Chemical Society*, vol. 118, pp. 3332–3339, 1996.

[31] T. Walton and J. W. Szostak, "A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension," *J. Am. Chem. Soc.*, vol. 138, no. 36, pp. 11996–12002, 2016.

[32] L. Li, N. Prywes, C. P. Tam, D. K. O Flaherty, V. S. Lelyveld, E. C. Izgu, A. Pal, and J. W. Szostak, "Enhanced nonenzymatic rna copying with 2-aminoimidazole activated nucleotides," *Journal of the American Chemical Society*, vol. 139, pp. 1810–1813, 2017.

[33] E. H. Ekland and D. P. Bartel, "RNA-catalysed RNA polymerization using nucleoside triphosphates," *Nature*, vol. 382, pp. 373–376, 1996.

[34] A. Wochner, J. Attwater, A. Coulson, and P. Holliger, "Ribozyme-catalyzed transcription of an active ribozyme," *Science*, vol. 332, no. 6026, pp. 209–212, 2011.

[35] S. Kauffman, *The origins of order: Self-organization and selection in evolution*. 1993.

[36] V. Vasas, E. Szathmry, and M. Santos, "Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 4, pp. 1470–1475, 2010.

[37] B. Rubinov, N. Wagner, M. Matmor, O. Regev, N. Ashkenasy, and G. Ashkenasy, "Transient fibril structures facilitating nonenzymatic self-replication," *ACS nano*, vol. 6, pp. 7893–7901, 2012.

[38] E. Guseva, R. N. Zuckermann, and K. A. Dill, "Foldamer hypothesis for the growth and sequencedifferentiation of prebiotic polymers," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 36, pp. E7460–E7468,

2010.

[39] J. Li, J. M. A. Carnall, M. C. A. Stuart, and S. Otto, "Hydrogel formation upon photoinduced covalent capture of macrocycle stacks from dynamic combinatorial libraries," *Angew. Chem., Int. Ed.*, vol. 919, no. 50, pp. 8384–8386, 2011.

[40] T. P. Fraccia, G. P. Smith, G. Zanchetta, E. M. Paraboschi, Y. Yi, D. M. Walba, G. Dieci, N. A. Clark, and T. Bellini, "Abiotic ligation of DNA oligomers templated by their liquid crystal ordering," *Nature Communications*, vol. 6, p. 6424, 2015.

[41] 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. U.S.A.*, vol. 109, no. 4, pp. 1110–1115, 2012.

[42] G. Zanchetta, T. Bellini, M. Nakata, and N. A. Clark, "Physical polymerization and liquid crystallization of RNA oligomers," *Journal of the American Chemical Society*, vol. 130, no. 39, pp. 12864–12865, 2008.

[43] N. Dolinnaya, T. A.V., V. Sergeev, T. Oretskaya, and Z. Shabarova, "Structural and kinetic aspects of chemical reactions in DNA duplexes. information on DNA local structure obtained from chemical ligation data," *Nucleic Acids Res.*, vol. 19, no. 11, pp. 3073–3080, 1991.

[44] N. Wrobel, M. Schinkinger, and V. Mirsky, "A novel ultraviolet assay for testing side reactions of carbodiimides," *Anal. Biochem.*, vol. 305, no. 2, pp. 135–138, 2002.

[45] A. F. Olea and J. K. Thomas, "Rate constants for reactions in viscous media: correlation between the viscosity of the solvent and the rate constant of the diffusion-controlled reactions," *J. Am. Chem. Soc.*, vol. 110, no. 14, pp. 4494–4502, 1988.

[46] M. I. Page and W. P. Jencks, "Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect,"

Proc. Natl. Acad. Sci. U.S.A., vol. 68, no. 8, pp. 1678–1683, 1971.

[47] C. Maffeo, B. Luan, and A. Aksimentiev, "End-to-end attraction of duplex DNA," *Nucleic Acids Res.*, vol. 40, no. 9, pp. 3812–3821, 2012.

[48] E. Horowitz, A. E. Engelhart, M. C. Chen, K. Quarles, M. W. Smith, D. G. Lynn, and N. V. Hud, "Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 12, pp. 5288–5293, 2010.

[49] M. Nowacka, P. Jackowiak, A. Rybarczyk, T. Magacz, P. M. Strozycki, J. Barciszewski, and M. Figlerowicz, "2D-PAGE as an effective method of rna degradome analysis," *Mol. Biol. Rep.*, vol. 39, pp. 139–146, 2012.

[50] A. Fradin, H. Gruhl, and H. Feldmann, "Mapping of yeast tRNAs by two-dimensional electrophoresis on polyacrylamide gels," *FEBS Lett.*, vol. 50, pp. 185–189, 1975.

[51] H. Gao, M. Yang, R. Patel, and A. F. Cook, "Circularization of oligonucleotides by disulfide bridge formation," *Nucleic Acids Res.*, vol. 23, pp. 2025–2029, 1995.

[52] D. Kowalski, "A procedure for the quantitation of relaxed closed circular dna in the presence of superhelical DNA: An improved fluorometric assay for nicking-closing enzyme," *Anal. Biochem.*, vol. 93, pp. 346–354, 1979.

[53] 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.*, vol. 107, no. 41, pp. 17497–17502, 2010.

[54] T. P. Fraccia, G. P. Smith, L. Bethge, G. Zanchetta, G. Nava, S. Klussmann, N. A. Clark, and T. Bellini, "Liquid crystal ordering and isotropic gelation in solutions of four-base-long DNA

oligomers," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, pp. 8508–8516, 2016.

[55] H. Franz, F. Ciuchi, G. Di Nicola, M. De Morais, and P. Mariani, "Unusual lyotropic polymorphism of deoxyguanosine-5'-monophosphate: X-ray diffraction analysis of the correlation between self-assembling and phase behaviour," *Phys. Rev. E*, vol. 50, pp. 395–402, 1994.

[56] G. C, L. Li, O. D. K., T. C.P., and J. W. Szostak, "A mechanistic explanation for the regioselectivity of nonenzymatic RNA primer extension," *J. Am. Chem. Soc.*, vol. 139, no. 46, pp. 16741–16747, 2017.

[57] P. Mariani, F. Spinozzi, F. Federiconi, H. Amenitsch, L. Spindler, and I. Drevensek-Olenik, "Small angle X-ray scattering analysis of deoxyguanosine 5-monophosphate self-assembing in solution: Nucleation and growth of g-quadruplexes," *J. Phys. Chem. B*, vol. 113, pp. 7934–7944, 2009.

[58] G. F. Joyce, T. Inoue, and L. E. Orgel, "Non-enzymatic template-directed synthesis on rna random copolymers. poly(C, U) templates," *J. Mol. Biol.*, vol. 176, no. 2, pp. 279–306, 1984.

[59] X. Sun, E. Cao, Y. He, and J. Qin, "Fluorescence studies on the interaction of ethidium bromide with duplex, triplex and quadruplex DNA structures," *China Ser. B-Chem.*, vol. 42, pp. 62–69, 1999.

[60] M. J. Cavaluzzi and P. N. Borer, "Revised UV extinction coefficients for nucleoside-5-monophosphates and unpaired DNA and RNA," *Nucleic Acids Res.*, vol. 32, no. 1, p. e13, 2004.

[61] J. Flory, "Molecular size distribution in linear condensation polymers," *J. Am. Chem. Soc.*, vol. 1, pp. 1877–1855, 1936.

[62] D. M. Soumpasis, "Theoretical analysis of fluorescence photobleaching recovery experiments," *Biophys. J.*, vol. 41, pp. 95–97, 1983.

[63] A. V. Tataurov, Y. You, and O. R., "Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids," *Biophys Chem.*, vol. 133, no. 1–3, pp. 66–70, 2008.

Appendices

2MeImpG reaction simulation

l = 1000; *number of elements*
time = 100000; *number of time steps for each simulation*

for round = 1:100 *number of simulation interations*
    j = ones(l,1); *array of 2MeImpG is initialized;*
    0 → rGMP, 1 → 2MeImpG or imidazolium-bridged dinucleotide;
    b = zeros(l,1); *array of chemical bonds is initialized;*
    0 → no bond, 1 → imidazolium-bridge bond, 2 → phosphodiester bond;

    for t = 0:time *simulation starts*
        n = randi([3 l-2]); *an element is randomly picked excluding the extremes;*
        direction = randi([0 1]); *a direction to propagate the reaction is chosen;*
        neigh = (direction-1+direction); *an array direction is defined to handle the indexes;*

        b(n) stores the chemical bond with j(n-1)
        b(n+1) stores the chemical bond with j(n+1)

        if (b(n + -(direction-1)) ~= 1 and b(n + direction) == 0 and j(n + neigh) == 1 and
        b(n + direction + neigh) == 1)
            b(n + direction) = 2;

```
                b(n + neigh + direction) = 0;
                j(n + neigh) = 0;
                j(n + neigh*2) = 1;


        elseif (b(n + -(direction-1)) ~= 1 and b(n + direction) == 0 and j(n +
neigh)== 1 and
            b(n + direction + neigh) ~= 1 and j(n) == 1)
                b(n + direction) = 1;


        elseif (b(n + -(direction-1)) == 1 and b(n + direction) == 0 and j(n + neigh)
== 1 and
            b(n + direction + neigh) == 1 and j(n) == 1)
                b(n + direction) = 2;
                b(n + neigh + direction) = 0;
                j(n + neigh) = 0;
                j(n + neigh*2) = 1;


        elseif (b(n + -(direction-1)) == 0 and b(n + direction) == 0 and j(n + neigh)
== 1 and
            b(n + direction + neigh) == 1 and j(n) == 1)
                b(n + direction) = 2;
                b(n + -(direction-1)) = 0;
                j(n + neigh) = 0;
                j(n + neigh*2) = 1;
        end
    end


    c = bwconncomp(b==2); elements connected by phosphodiester bond are re-
covered;
    for i = 1:c.NumObjects
        products(i) = size(c.PixelIdxListi,1);
    end
```

b($b==1$) = 0; c_{mon} = bwconncomp($b==0$); *elements not connected by phosphodiester bond are recovered;*

    for i = 1:c_{mon}.NumObjects

        mons(i) = size(c_{mon}.PixelIdxListi,1);

    end

    mons = mons-1;

    h(:,round) = histcounts(products+1,[1:10]); *products size distribution for each round;*

    mon(round) = sum(mons);

end

hh = mean(h,2).*[1:9]'./(l-4);

mon_h = mean(mon);

hh_{mol} = mean(h,2);

hh_{mol}(1) = mean(mon);

hh(1) = mon_h./(l-4);

hh_{mol} = hh_{mol}./sum(hh_{mol});

Paper 1

Nonenzymatic Polymerization into Long Linear RNA Templated by Liquid Crystal Self-Assembly.

I have contributed to this work as follows:

– Experimental design and paper writing;

– Phase diagram characterization of 12BE;

– Phase diagram characterization of 12BE + PEG8k;

– Ligation experiments for 12BE and partially for 6SE;

– PAGE for 12BE reaction products and partially for 6SE;

– Analysis of 12BE ligation products;

– 2D PAGE experiments for 12BE ligation products;

– FRAP measurements for Fluorescein diffusion;

– Study of the differential solubility of small solutes in iso-iso mixtures;

– Kinetic measurements for EDC-mediated ligation;

– Kinetic modeling with the supervision of Tommaso Bellini;

– 6SE enzymatic digestion and products analysis.

ARTICLE

Nonenzymatic Polymerization into Long Linear RNA Templated by Liquid Crystal Self-Assembly

Marco Todisco,[†] Tommaso P. Fraccia,[†,‡] Greg P. Smith,[§] Andrea Corno,[†] Lucas Bethge,[⊥] Sven Klussmann,[⊥] Elvezia M. Paraboschi,[¶] Rosanna Asselta,[¶,#] Diego Colombo,[†] Giuliano Zanchetta,[†] Noel A. Clark,[§] and Tommaso Bellini*,[†]

[†]Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università di Milano, via Vanvitelli 32, 20129 Milano, Italy

[‡]Dipartimento di Scienze Umane e Promozione della Qualità della Vita, Università San Raffaele di Roma, via di Val Cannuta, 247, I-00166 Roma, Italy

[§]Department of Physics and Soft Materials Research Center, University of Colorado, Boulder, Colorado 80309-0390, United States

[⊥]NOXXON Pharma AG, 10589 Berlin, Germany

[¶]Department of Biomedical Sciences, Humanitas University, via Rita Levi Montalcini 4, Pieve Emanuele, Milano I-20090, Italy

[#]Humanitas Clinical and Research Center, via Alessandro Manzoni 56, Rozzano, Milano I-20089, Italy

Ⓢ *Supporting Information*

ABSTRACT: Self-synthesizing materials, in which supramolecular structuring enhances the formation of new molecules that participate to the process, represent an intriguing notion to account for the first appearance of biomolecules in an abiotic Earth. We present here a study of the abiotic formation of interchain phosphodiester bonds in solutions of short RNA oligomers in various states of supramolecular arrangement and their reaction kinetics. We found a spectrum of conditions in which RNA oligomers self-assemble and phase separate into highly concentrated ordered fluid liquid crystal (LC) microdomains. We show that such supramolecular state provides a template guiding their ligation into hundred-bases long chains. The quantitative analysis presented here demonstrates that nucleic acid LC boosts the rate of end-to-end ligation and suppresses the formation of the otherwise dominant cyclic oligomers. These results strengthen the concept of supramolecular ordering as an efficient pathway toward the emergence of the RNA World in the primordial Earth.

KEYWORDS: *self-assembly, liquid crystals, supramolecular chemistry, RNA World, origin of life, ligation, microreactors*
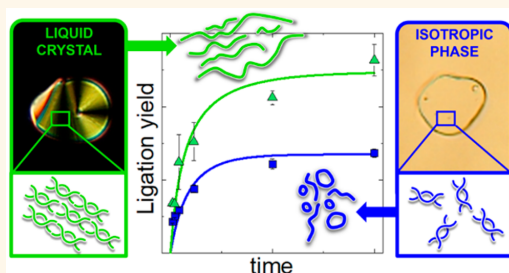
The growing awareness of the relevance of supramolecular ordering on chemical reactivity suggests new routes for the emerging of patterns and structures.[1−4] Relevant examples are given by the self-synthesizing materials, in which self-assembly of amphiphilic molecules provides a template for covalent bond formation, which in turn promotes and stabilizes the supramolecular structures.[5,6] Indeed, it has been observed that supramolecular assembly can play a pivotal role in chemical reactions, directing the regioselectivity and boosting the reaction kinetics. Through this route, new symmetries and patterns may arise, reflecting more the mechanisms of molecular assembly than the reactivity of the isolated substrate. Because of this, it has been suggested that self-assembly guided reactions could emerge as a rich strategy for

the design of nanostructured materials and as a promising pathway to understand fundamental mysteries such as the origin of life.[1,6−11] In this last context, one of the main challenges is understanding the spontaneous formation of nucleic acid polymers from their monomeric constituents[12,13] since it has to overcome several critical issues, such as the extreme low reactivity of nucleotides,[14,15] the spontaneous tendency to form cyclical products contrasting the elongation of the polymers,[16,17] the high hydrolysis rate of phosphodiester

bonds in aqueous environments[18] especially for RNA, and the competition of a variety of other potentially available reactive species, which would require a powerful selection mechanism.

In this study, we demonstrate and quantitatively characterize a nonenzymatic reaction pathway based on supramolecular assembly that favors the formation of long linear RNA chains by ligating short RNA oligomers. This work stems from a previous observation that the onset of liquid crystal ordering in solutions of DNA 12mers promotes their polymerization.[19] By a combination of different approaches, we demonstrate here the templating effect of the self-assembly of RNA oligomers, by which the geometry of the supramolecular aggregates favors intermolecular reactions over intramolecular reactions, while the local packing boosts the reaction rate. In this study, we considered shorter (6-base-long) and longer (12-base-long) RNA oligomers forming duplexes with terminal overhangs or blunt-ends, respectively, and we considered natural D-RNA and enantiomeric L-RNA.

LIQUID CRYSTAL ORDERING AS A GENERAL SUPRAMOLECULAR MOTIF OF NUCLEIC ACIDS

The reversible but selective interaction among nucleic acid chains at the core of the processing of genetic information relies on pairing and stacking interactions of nucleobases and on the good water solubility of the chains, granted by the ionized phosphate groups. These features confer to nucleic acids rather unique self-assembly properties, enabling the formation of extremely complex natural and artificial structures, such as the hierarchical structure of chromatin and the multistrand constructs of DNA nanotechnology.[20] In a series of papers,[21–26] we have shown that the same structural features and interactions lead, in aqueous solutions of short DNA oligomers at high concentration (c_{DNA} > 200 mg/mL), to long-ranged supramolecular liquid crystal (LC) ordering afforded by hierarchical steps of self-assembly. LC ordering appears indeed as a rather general form of spontaneous ordering in DNA solutions and can be accompanied by the segregation of DNA LC domains when duplex-forming oligomers are mixed with unduplexed DNA strands, or with flexible polymers (e.g., PEG),[27] or with duplexes incapable to form linear columns.[23]

LIQUID CRYSTAL AS A STEP FOR THE ORIGIN OF LIFE

The emergence of early life and DNA as the genetic polymer is widely thought to have proceeded through the so-called "RNA World", featuring RNA filaments playing the double role of carrying information and sustaining primordial metabolism.[28,29] Such catalytic RNA strands, the "ribozymes", first discovered by Cech and co-workers,[30] are chains having specific sequences and secondary structures. Tube-test evolution indicates that the minimum length of an RNA strand capable of providing the simplest enzymatic activity (cleavage of an RNA substrate) is of only 16 nt,[31] while more complex catalytic activities, such as the template-directed polymerization of NTPs on an external template, require much longer sequences, in the order of 150 nt.[32] These are challenging lengths to be reached starting from mononucleotides or even from short oligomers since their spontaneous formation is highly improbable. Indeed, the origin of the RNA World[33] is one of the main unsolved enigma of our understanding of nature.

In the search of a convincing "guiding hand", capable of selecting and ligating simple molecular groups into RNA strands,[8,34] various mechanisms have been proposed relying on templates and enhancers such as clay surfaces,[35,36] lipid assemblies,[37–40] and pairing of nicked oligomers on a templating RNA strand.[41,42] The crucial search of a prebiotically plausible leaving group for phosphate condensation reaction has fostered the study of different compounds based on imidazole to overcome the low reactivity of polyphosphate.[43,44] The investigation of efficient methods to enhance linear polymerization over cyclization, a must to obtain long RNA strands, has highlighted the importance of base pairing and stacking interactions of building blocks[45] and the role of aromatic base-intercalating molecules.[16]

On the basis of the intimate connection between features that enable propagating genetic information, and those that promote LC ordering, we recently proposed[19,46] that the LC ordering of ultrashort nucleic acid oligomers could provide a positive feedback loop driving their abiotic ligation into longer oligomers. Indeed, in condensed LC phases, the oligomers are stacked and organized into columns, each being formed by a physically continuous but chemically discontinuous double helix in which the oligomer termini are held in close proximity. According to this scenario, the supramolecular assembly acts as a template guiding intermolecular ligation toward the formation of long linear chains when in the presence of appropriate reaction conditions. At the same time, the formation and stability of LCs is strongly dependent on the molecular shape anisotropy, since the presence of longer double helices lowers the concentration required for the onset of LC ordering. This establishes a positive feedback by which the ligation reaction guided by liquid crystals ordering promotes their own self-assembly.

In previous experiments performed in solutions of 5′(OH)−3′(p) DNA 12-base-long oligomers, we found that the onset of LC ordering markedly increased the interoligomer polymerization yield.[19]

On the basis of this seminal study, we decided to explore the intimate nature of the templating effect of the liquid crystal ordering of nucleic acids in terms of reaction rates and products structure, as a nonpeptidic example of self-synthesizing material. We focus here on RNA molecules, generally considered a more ancient nucleic acid than DNA. The extension of the LC properties from DNA to RNA is not trivial since RNA has an A-form duplex helical structure, with paired bases tilted with respect to the helical axis,[47] a geometry of stacking that could affect the supramolecular structure.[48] While the conditions at which the prebiotic synthesis could have occurred are basically unknown, current scenarios are based on the notion of drying-wetting and heating−cooling cycles, possibly in the presence of thermal and chemical gradients such as those that can be found in inland hydrothermal springs[49] or in porous minerals surrounding oceanic hydrothermal vents.[50] The experiments we describe here are in conditions of mild temperature, and of molecular concentrations and ionic strengths naturally achieved in drying cycles, and thus in line with the current explorations in the research devoted to the origin of life.

RESULTS AND DISCUSSION

RNA Oligomers. We studied two RNA oligomers of length 12 and 6:

(i) 5′-pCGCGAAUUCGCG-3′ (12BE), a self-complementary sequence forming terminal blunt-ends whose DNA analog we previously investigated.[21] 12BE used in this

study is L-RNA, synthesized using L-ribose, and its structure is thus mirror symmetric with respect to natural RNA.[51] The use of L-RNA has the considerable advantage of being unaffected by the action of natural RNase, which has the potential of significantly affecting the quantitative analysis here proposed.

(ii) 5′-pGAUCGC-3′ (6SE), a partially self-complementary sequence. 6SE form 4-base-long paired duplexes with 2-base-long GC "adhesive" overhangs. 6SE in this study is D-RNA and is thus potentially affected by RNase degradation. We indeed used enzymatic degradation as a tool to further characterize 6SE LC-promoted RNA ligation.

Both 12BE and 6SE terminate, on the 5′ end, with a phosphate group that provides the starting point of the ligation reaction. We opted for the 5′ terminus instead of the 3′ end to make sure that no cyclization between 2′ and 3′ on the same sugar moiety could occur. With this choice, the ligation could lead to either 5′-2′ or 5′-3′ phosphodiester bonds. This depends on the regioselectivity in the nonenzymatic formation of phosphodiester bonds, which, in the case of interduplex ligation, is unknown. The ligation of nicked oligomers on a template strand has been found to be highly selective toward 5′-3′ linkage,[52,53] despite the fact that 2′-hydroxyl group is more nucleophilic than 3′-hydroxyl group.[54]

Abiotic Activator. Even if it is not considered a prebiotically plausible molecule, as proof of concept, we performed condensation reactions by using the water-soluble carbodiimide EDC (N-ethyl-N′-(dimethylaminopropyl) carbodiimide), which can react with terminal phosphate groups making them reactive toward nucleophiles, like primary amines[45] (yielding phosphoramidates), hydroxyl NA termini[55] (phosphodiesters), or water (hydrolysis). Surprising high yields (more than 90%) were previously observed in DNA LC,[19] which suggest a strong templating effect of this form of supramolecular ordering.

Phase Behavior. RNA oligomers spontaneously order into LC phases. Figure 1 illustrates typical textures observed in polarized optical microscopy showing coexistence of chiral nematic (N*) and columnar (COL) phases (panel a, showing a COL domain surrounded by N*) and COL phase (panels b and c, where the COL phase is recognizable from the characteristic fan-shaped domain). These phases appear at high RNA concentrations ($c_{RNA} > 250$ mg/mL at room temperature) and in a broad interval of temperatures (T), as shown in the phase diagrams of 12BE (Figure 1d) and 6SE (Figure 1e). For both systems, the solutions display N* ordering (blue shading) and COL ordering at increasing c_{RNA} (red shading). Overall, the two phase diagrams are rather similar, indicating that the smaller axial ratio of 6SE duplexes, typically disfavoring LC ordering,[21] is compensated by their stronger overhang-mediated end-to-end coupling. The phase boundaries reported here for 12BE are not significantly dissimilar from those found for the analogous DNA molecule. These results suggest that the RNA interduplex stacking is at least as large as in the case of DNA.

To identify the best conditions for ligation experiments, we underwent a more detailed analysis of the phase behavior. First, we tested the effects of Mg^{2+}, which we found to stabilize the RNA LC phases. This is shown by the dashed purple lines in the phase diagrams that mark the LC-isotropic phase boundary when 12BE and 6SE solutions are prepared with

60 mM $MgCl_2$. We then studied the phase behavior of mixtures of 12BE and PEG (MW 8000), employed as molecular crowding agent. We previously found that for sufficiently large PEG concentration, DNA/PEG solutions demix into coexisting DNA-rich and PEG-rich fluids.[19,27] We thus made an effort to determine the conditions under which the system phase separates adopting the morphology of RNA-rich microdomains surrounded by the PEG-rich isotropic phase. In this condition, the system is effectively formed by microreactors where the reaction takes place, which are held by PEG osmotic pressure with no membrane or barrier for the diffusion of ions, EDC and waste products.

We thus carefully explored the phase behavior obtained by solubilizing a moderate amount of 12BE ($c_{RNA} \approx 37$ mg/mL) as a function of PEG concentration, as shown in Figure 1f. The solutions are prepared in 60 mM $MgCl_2$ to take advantage of the stabilizing effect of Mg^{2+} ions. We find that for $c_{PEG} \leq 100$ mg/mL, the solution remains homogeneous, while for $c_{PEG} \geq 150$ mg/mL, RNA-rich microdomains start to appear. Depending on T, the RNA-rich domains are found either in isotropic (ISO) state (orange dots, orange shading) or as LC (green triangles, green shading). The two conditions are shown in the polarized microscopy pictures, where liquid–liquid (Figure 1g) and liquid–LC (Figure 1h) phase coexistence can be distinguished by the colors appearing in the birefringent LC domains as opposed to the dark, isotropic PEG-rich background. By a quantitative analysis of the phase diagram (described in the Supporting Information), we determined the limiting solubility of RNA in the PEG-rich phase to be $c_{RNA} \approx 13$ mg/mL, and the RNA concentration within the microdomains for two PEG concentrations: {$c_{PEG} = 150$ mg/mL, $c_{RNA} \approx 230$ mg/mL} and {$c_{PEG} = 200$ mg/mL, $c_{RNA} \approx 600$ mg/mL}. This analysis enabled us to pinpoint the conditions in which well-determined amount of RNA oligomers are effectively confined in membrane-less microreactors. We also found that an analogous phase separation and confinement effect can be obtained by mixing 12BE with single strand RNA (sequence 5′-CCUCAAAACUCC-3′), as described in the Supporting Information.

The phase diagram in Figure 1f enables identifying T = 40 °C as a convenient isothermal condition where to investigate the efficiency of EDC promoted ligation in mixtures of 12BE and PEG. At this T, upon increasing c_{PEG}, the system transitions from homogeneous mixture (where RNA is diluted, $c_{RNA} \approx 37$ mg/mL), to liquid–liquid coexistence (where RNA is concentrated but disordered, $c_{RNA} \approx 230$ mg/mL) to liquid–LC coexistence (where RNA is concentrated and ordered, $c_{RNA} \approx 600$ mg/mL).

A similar condition could not be found for 6SE/PEG mixtures since we find liquid–LC coexistence at room T (T = 25 °C) for $c_{PEG} \geq 100$ mg/mL, but no liquid–liquid coexistence at any T since upon heating the system transforms directly from liquid–LC coexistence to homogeneous (Figure S17). We thus explored EDC promoted ligation in 6SE/PEG mixtures at room T.

Abiotic Ligation in Solutions of RNA Oligomers. Figure 2 summarizes the results of ligation experiments in 6SE/PEG ($c_{RNA} \approx 100$ mg/mL ≈ 50 mM) and in 12BE/PEG ($c_{RNA} \approx 37$ mg/mL ≈ 10 mM) mixtures with 60 mM $MgCl_2$, 0.01 M HEPES buffer pH 7.5, molar ratio EDC/RNA of 30, and various c_{PEG}. The energy source is provided by the water-soluble condensing agent EDC, through the reaction summarized in Figure 4a. The ligation products have been
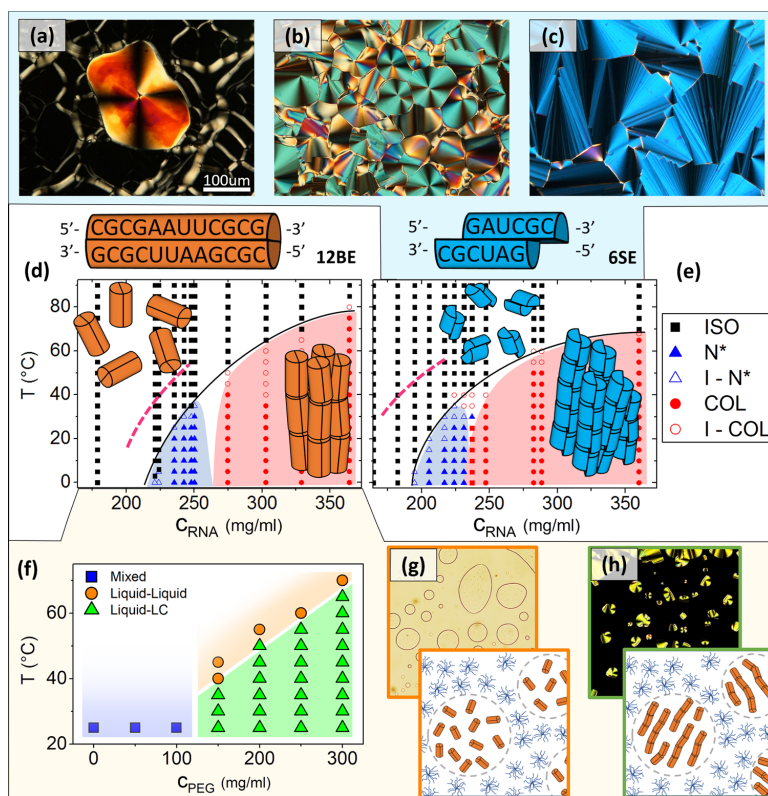
Figure 1. Supramolecular liquid crystal ordering in aqueous solutions of RNA oligomers. (a, b, c) Pictures by polarized optical microscopy of liquid crystalline phases in 6mer 6SE: (a) nematic/columnar coexistence and columnar phases (b, c_{RNA} = 250 mg/mL; c, c_{RNA} = 350 mg/mL). Size bar shown in panel a is shared by b and c. (d, e) Concentration (c_{RNA})−temperature (T) phase diagrams of the RNA (d) 12mer 12BE and (e) 6mer 6SE, whose duplex structures and self-assembly are sketched. Blue triangles and shading: chiral nematic phase (N*). Red dots and shading: columnar phase (COL). Black squares: isotropic phase (ISO). Empty dots: ISO-LC coexistence. Pink dashed line marks where the N* phase boundary is found when 60 mM Mg^{2+} is added to the solution. (f) PEG8k concentration (c_{PEG})−T phase diagram for 10 mM 12BE, 60 mM $MgCl_2$, yielding homogeneous phase (blue squares and shading), liquid−liquid coexistence (orange dots and shading, and orange-framed microscope picture and sketch, panel g), liquid−LC coexistence (green triangles and shading, and green-framed polarized microscope pictures and sketch, panel h).

evaluated by stopping the reaction at various time points (from 15 min to 48 h) by 20-fold dilution with 50 mM ethanolamine and by analyzing the mixture by 15% polyacrylamide gel electrophoresis in denaturing conditions, 7 M urea, and staining with SYBR Gold (see Supporting Information). The phase separation and LC nucleation time is much faster than all reaction times here explored. The outcome of 48 h reactions is shown in Figure 2a for 12BE/PEG, each gel lane corresponding to different c_{PEG} and phases, as also described by the associated pictorial sketches. The presence of longer chains is enhanced by the phase separation (orange frame) and markedly boosted by the appearance of LC ordering (green frame).

Examples of profiles of the fluorescent intensity in the gels $I_F(x)$ are shown in Figure 2c (black lines) for the ligation experiments with c_{PEG} = 0 mg/mL and c_{PEG} = 250 mg/mL, where the x coordinate along the lanes provides the product length associated with the bands. A reasonable estimate of the polymerization yield, defined as the fraction of newly formed phosphodiester bonds over the starting available 5′ terminal

phosphates, can be obtained by comparing the experimental cumulative I_F (colored dots) with the one predicted by a Flory model for simple polymerization (continuous line),[19,56] indicating that LC ordering leads to a markedly larger yield. Equivalently, the mean size of the products $\overline{n}_b$, plotted in Figure 2d as a function of the PEG concentration shows that the increment associated with LC ordering is evident. According to this analysis in LC phase, we obtain polymerization yields of 66%, 71%, and 69% for c_{PEG} = 200 mg/mL, 250 mg/mL, and 300 mg/mL, respectively, corresponding to a fraction of more than 50% of the products being more than five-times larger than the monomer ($n_b \geq 60$). This finding contrasts the moderate ligation yields found in the mixed solutions and in the liquid−liquid coexistence, where the average length of the products never exceeds twice and three-times the monomer length, respectively.

Similar results are found for 6SE/PEG (Figure 2b), where the enhancement of ligation efficiency brought about by phase separation and LC ordering is evident. In this case, though, the
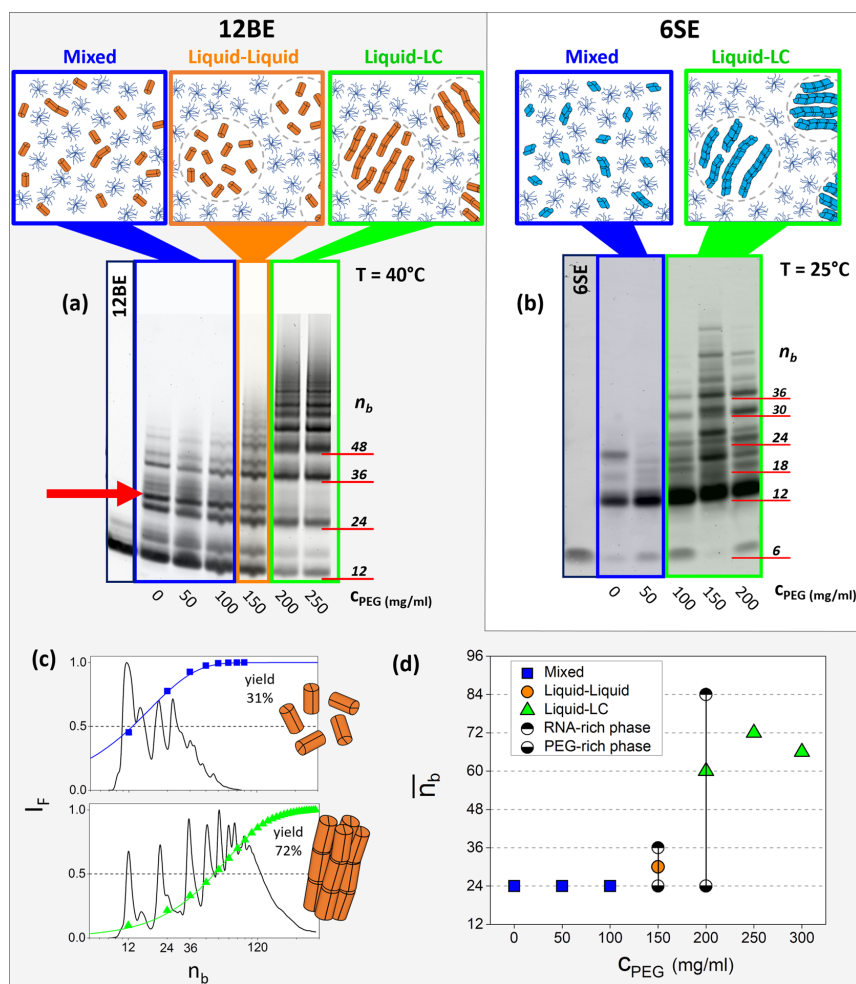
Figure 2. Ligation reaction in mixtures of RNA/PEG8k/MgCl$_2$/EDC. (a, b) Denaturing 15% PAGE of ligation products after 48 h of reaction of 12BE at T = 40 °C (a) and 6SE at T = 25 °C (b) for various PEG8k content (c_{PEG}). Blue, orange, green frames and sketches indicate the mixed state, liquid–liquid coexistence and liquid–LC coexistence, respectively, in which reactions are performed. Lengths expressed as number of bases (n_b) are marked on the right-hand side of the PAGE gel pictures. The red arrow marks a double band discussed in the main text. (c, d) Analysis of the reaction products for 12BE. (c) Fluorescence intensity (IF) profiles of the c_{PEG} = 0 mg/mL and c_{PEG} = 250 mg/mL lanes (black line) plotted versus n_b. Integrated IF (symbols) and their best fit to the Flory simple polymerization model (colored lines). (d) Average n_b of the reaction products versus c_{PEG} (symbols and color code refer to the phase in which reactions are performed). For c_{PEG} = 150 mg/mL and c_{PEG} = 200 mg/mL, the average n_b has been separated into the contributions from the two coexisting (RNA-rich and PEG-rich) phases.

product mass distribution is not well described by the Flory model due to a lower fluorescence emission of the monomer. This limitation prevents us to assess the reaction yield via gel analysis. However, the low molecular weight of 6SE enabled us to perform nuclear magnetic resonance (NMR) and unambiguously assign the signals in the ^{1}H and ^{13}C spectra. Particularly, ^{1}H NMR reveals broadening of the peaks in the 6SE sample treated with EDC (Figure 3a), a well-known feature reflecting polymer formation.[57] ^{13}C NMR experiment allows the study of the anomeric carbons of the nucleotides sugars. As apparent in Figure 3b, EDC treatment causes the

disappearance of the signal assigned to the first nucleotide of the monomer (1′ G1) and the appearance of a new signal assigned to the first nucleotide (1′ G8) of the molecule bound to another oligo, thus offering an additional evidence of ligation. The quantitative comparison of the G8 peak with a reference peak (i.e., the peak at 88.5 ppm remains unchanged between unreacted and reacted sample) and with the background noise indicates that in the ligation reaction of 6SE occurring in LC, the fraction of 5′P termini involved in phosphodiester bonds is of the order of 90% (see Full NMR Spectra file and Supporting Information for more details).
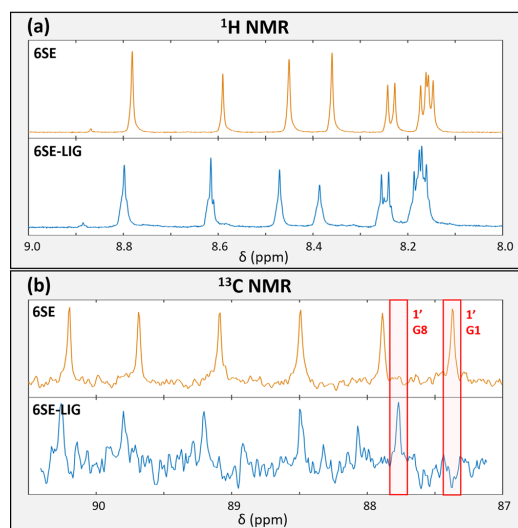
Figure 3. NMR study of 6SE ligation in LC state. (a) Comparison of the aromatic bases protons region of ^{1}H NMR spectra of 6SE (top panel) and 6SE-LIG at 343 K (bottom panel). Reacted samples show broadening of the peaks, giving a first clue for the formation of polymers in solution. (b) Comparison of ^{13}C NMR spectra of 6SE (top panel) and 6SE-LIG at 343 K (bottom panel). The peak at 87.4 ppm (1′ G1), corresponding to the anomeric carbon of the first G nucleotide of the monomer, disappears upon EDC treatment, while a peak at 87.8 ppm (1′ G8) appears. This second peak is assigned to the anomeric carbon of the first G nucleotide belonging to a molecule preceded by a bound oligo, thus demonstrating ligation. Full NMR spectra are reported in the Full NMR Spectra.

In principle, LC-promoted RNA ligation could also influence the regioselectivity (5′-3′ vs 5′-2′) of the newly produced phosphodiester bonds. The NMR analysis of 6SE ligation products (see Supporting Information) reveals no peak corresponding to 2′-5′ phosphodiester bond formation, as expected for complementary strand templating,[52,53] making it hard to decouple a possible contribution of the supramolecular ordering. At the moment, our data do not enable assessing the role of LC-promoted RNA ligation alone in guiding the regioselectivity of the products, which could in principle be extracted by 12BE ligation products analysis.

Ligation Reaction Kinetics: Dependence on Supramolecular Ordering. In the abiotic ligation reaction scheme exploited in this work (Figure 4a), the energy source driving the reaction is provided by EDC, that can either hydrolyze into water (with a rate k_h) or activate the phosphate termini (k_1) producing an intermediate (Figure 4a, red frame), which could in turn hydrolyze into EDU (N-ethyl-N'-(3-(dimethylamino)-propyl) urea) and a phosphate terminus (k_2), or produce a phosphodiester bond with the 3′ or 2′ hydroxyl terminus (k_L). Our data point to an enhancement of ligation efficiency linked to the transition to liquid crystalline state (Figure 2), which could be due to a loss of entropy of activation thanks to the positional constraint provided by the LC-stabilized columnar stacking of duplexes that holds the duplex termini in continuous contact. Since this enhancement effect could in principle also originate from the higher RNA concentration of

the LC phase, or even take place outside the LC domains, we performed an extended set of measurements aiming at quantifying the reaction rates to clarify whether LC ordering can indeed be considered a templating milieu. The analysis outlined below and detailed in the Supporting Information indicates that the effect of LC ordering can be described as an enhancement of k_L by at least six-times.

In a first group of observations, we measured the evolution in time of the absorption spectra of solutions of either EDC only, or EDC with various concentrations of 5′phosphate ribose. This molecule, in which the phosphate moiety is analogous to the 5′ termini of oligonucleotide chains, has been chosen to study the activation of the phosphate because of its negligible absorbance in the UV range of interest (Figure 4b). These measurements were performed in solutions of 60 mM $MgCl_2$, and HEPES buffer pH = 7.3, adjusted to match the pH measured in LC-forming PEG/EDC/RNA solutions. The measured spectra can all be perfectly fitted by a combination of the EDC and EDU spectra (Figure 4d), with amplitudes that evolve in time as exponentials and sum up to a time-independent constant (Figure 4e), as visually noticeable by the neat isosbestic point. No trace of the absorption spectrum of the activated phosphate could be observed, despite its chemical structure would suggest a detectable absorbance, an indication that the concentration of this group is negligible, in turn indicating a large k_2 value. By analyzing the data with a set of coupled linear kinetic equation describing the reaction steps in Figure 4a, we could exploit this simplified condition to quantify the rates k_h and k_1 (see Supporting Information), while for k_2 only lower boundaries can be given.

To pinpoint the effects of LC ordering on ligation, we studied in depth the ligation reaction in the PEG/EDC/12BE mixture by starting from the simplest condition, that is, at low PEG (c_{PEG} = 100 mg/mL), where the system remains in the mixed state. We examined the ligation products obtained in this condition as a function of the reaction time (blue squares in Figure 4c) and fitted to the same set of coupled linear reaction equations. By keeping k_h and k_1 as determined above, the model predicts the yield to depend only on the ratio k_L/k_2. By fitting the data, we obtain a nicely approximating curve (blue line in Figure 4c) and $k_L/k_2 \approx 100$ (Supporting Information), indicating that the activated terminus is more reactive with the ribose OH at 2′ or 3′ position than with water.

At higher PEG concentration, where the system phase separates, the reaction becomes more complex since it could take place simultaneously in the two coexisting phases. This condition can still be modeled by a larger set of coupled linear equations (see Supporting Information), whose parameters include the relative volumes filled by the two phases and the RNA concentrations in both phases. Being diffusion of reagents an essential ingredient of the process, we made sure that the condensed phase was fluid and that EDC and EDU could diffuse in and out the domains. This was done by FRAP (fluorescence recovery after photobleaching) measurements, which indicate that the equilibration time of molecules of the size of EDC and EDU takes place in a time much shorter than the reaction time so that reaction-induced local EDC depletion can be neglected (see Supporting Information). We assume k_h and k_2, rates of reaction with water molecules, to be the same as in the mixed condition since hydration always remains significant. Conversely, k_1 and k_L might be, in the RNA-rich phase, significantly different than in the mixed state because of two concurring effects. First, the solubility of EDC in the RNA-
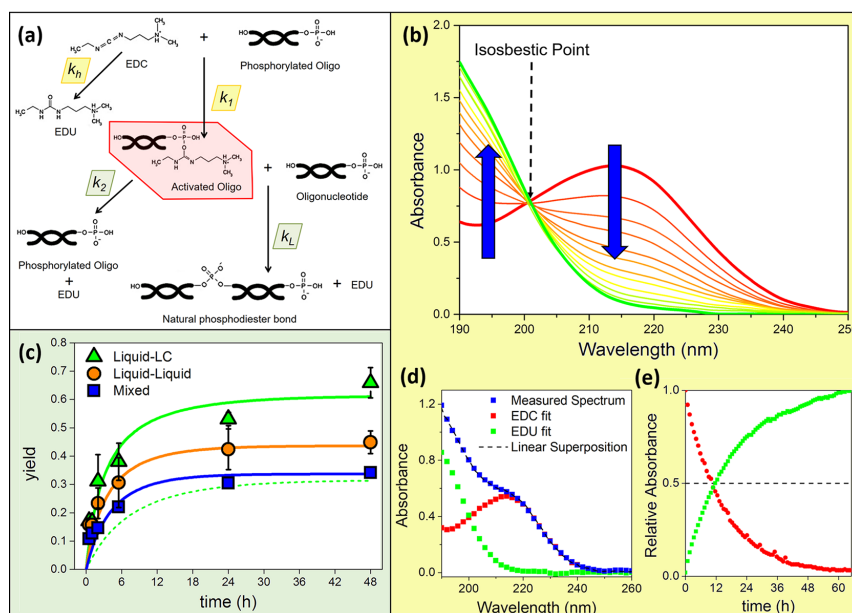
Figure 4. Ligation reaction kinetics. (a) Reaction steps and definition of the reaction rates. The red frame identifies the activated intermediate. Parameters shaded in yellow and green are determined as described in panels b, d, e, and c, respectively. (b) Example of k_1 measurement through UV-spectra acquisition over time for a diluted EDC + 5′phosphate ribose (P) mixture. Arrows indicate the evolution over time of the spectra, that show an isosbestic point. (d) Decomposition of the measured spectra as linear superposition of the EDC and EDU spectra. The relative amount of the two species is plot over time as shown in (e) and fit to determine k_1 and k_h. (c) 12BE polymerization yield versus time in 10 mM RNA/PEG8k/60 mM MgCl$_2$/300 mM EDC in mixed solutions (c_{PEG} = 100 mg/mL, blue squares), liquid−liquid coexistence (c_{PEG} = 150 mg/mL, orange dots), liquid−LC coexistence (c_{PEG} = 200 mg/mL, green triangles) at T = 40 °C. Continuous lines are the best fit obtained by modeling the reaction with a set of coupled linear equation. Fit to the mixed state (blue line) and to the liquid−liquid coexistence (orange line) enable determining all kinetic parameters. Dashed green line: prediction of the yield versus time for the liquid−LC coexistence by this choice of parameters. Continuous green line: prediction by the linear kinetic model when k_L is increased by six-times.

rich solutions might be much smaller than in the PEG-rich phase, as suggested by the strong partitioning observed for fluorescent dopants of analogous molecular size (see Supporting Information). Moreover, the viscosity of the RNA-rich phase markedly grows with RNA concentration, as indicated by FRAP measurements. Since reaction rates are typically inversely proportional to viscosity[58] and viscosity grows more than linearly with RNA concentration (see Supporting Information), we expect k_1 and k_L to be decreased at least proportionally to the c_{RNA}: $k_1' = (F/c_{RNA}) \, k_1$; $k_L' = (F/c_{RNA}) \, k_L$, where F is proportionality coefficient. Ligation kinetics in liquid−liquid coexistence (c_{PEG} = 150 mg/mL, orange circles in Figure 4c) is well approximated by the model with the choice F = 30 mM (orange line).

When we apply the same procedure with the same choice of rates k_h, k_1', k_2, k_L', to model the ligation kinetics in liquid−LC coexistence (green triangles in Figure 4c), the model strongly underestimates the observed yield (dashed green line). To match the data we need to increase k_L of at least six-times, a choice that leads to the continuous green line in Figure 4c. In other words, we find that the rate of interstrand ligation in LC is enhanced over what expected in an identical system that had no LC ordering by six-times, possibly much more, indicating that the LC ordering is indeed promoting ligation, con-

firming the templating effect of this supramolecular self-assembly. We emphasize that the six increment is a lower estimate since our FRAP experiments indicate that viscosity increases much more than proportionally to concentration as the LC ordering appears.

Having a tool to evaluate the EDC-promoted ligation in phase separated PEG/EDC/RNA solutions, we can split the observed yield into the contributions from the two coexisting phases (Figure 2d, half dots). The model with the parameters determined as above indicates that the largest majority of the longer strands are indeed produced by the ligation reaction in the LC phase. The total yield, evaluated by gel electrophoresis, is effectively lowered by the inefficient reaction in the PEG-rich, RNA-poor phase that fills most of the volume (see the determination of concentrations in RNA-PEG separate phases in Supporting Information).

Ligation Products: Dependence on Supramolecular Ordering. By holding duplexes in a fluid but ordered matrix, LC ordering not only promotes the formation of chemical bonds between the contacting termini, but also inhibits intramolecular ligation. The molecular packing within the LC phase stabilizes the duplexes and holds them stacked into linear aggregates thus disfavoring the contacts between the termini within each duplex. Circular structures are instead expected to
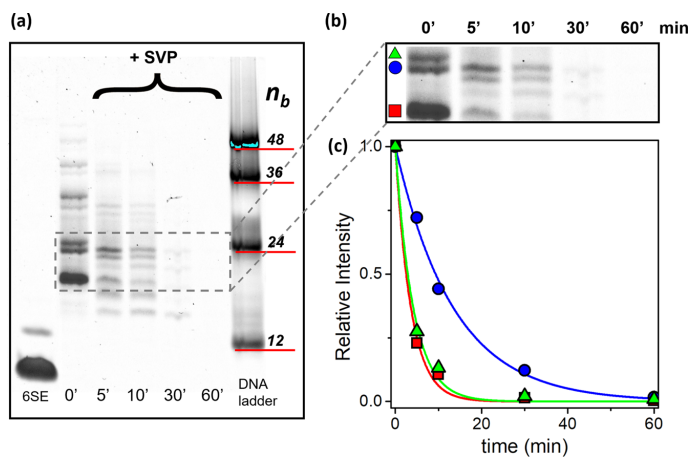
Figure 5. Discrimination of 6SE linear versus circular ligation products by SVP digestion. (a) Denaturing 20% PAGE of 6SE reaction products in liquid−LC coexistence (c_{PEG} = 250 mg/mL) treated with Snake Venom Phosphodiesterase I (SVP). The digestion reaction has been stopped at different times (5′, 10′, 30′, 60′) to point out bands more efficiently digested. Gel contrast has been enhanced for a better visualization. (b) Highlight of bands associated with lower weight products and (c) plot of relative intensities of the bands over time. Continuous lines are guides for the eye. An alternate pattern of easily and hardly digestible products emerges in the PAGE band sequence.

be a favorite outcome of the ligation reaction of duplexes in isotropic phases, where the transient melting of the terminal bases and their reaction can easily occur. This notion is supported by two distinct approaches described here: selective enzymatic digestion of 6SE ligation products and 2D-PAGE of 12BE products, which, being L-RNA, cannot be enzymatically probed. This conclusion is further supported by NMR data, as discussed below.

In PEG/EDC/12BE mixtures, reaction products obtained in disordered solutions (Figure 2a, blue and orange frames), contain a larger variety of smaller oligomers than reaction products in the LC phase (green frame). This is the case of the product leading to the PAGE band found in between the linear 24mer and 36mer and marked in Figure 2a by a red arrow, which becomes almost undetectable in LC reactions. To explore the difference between these bands, we have run 2D-PAGE experiments, an older technique to characterize mixtures of RNA molecules that are not resolved in a simple 1D-PAGE experiment. This technique, which takes advantage of the fact that the migration velocity of oligomers of distinct topology depends differently on the concentration of the gel,[59] has been used to characterize complex tRNA mixtures and solutions of RNA polymers.[60] By using 2D-PAGE, we find a markedly different behavior of the red arrow-marked band with respect to the one closest to it, indicating that they have distinct topologies (see Supporting Information). The simplest topological difference between 24mers is linear versus circular. Indeed, circular ssRNAs are known to have a slower electrophoretic mobility as compared to the linear molecules of equal length,[61] in line with the position of this extra band of different mobility.

Production of circular oligomers also appears as the reason behind the lower apparent ligation yield in 6SE PAGE. The two-base long overhangs make the formation of phospho-diester bonds between the 5′ and 3′ (or 2′) termini of unbound 6SE duplexes much easier than in the blunt-ended structure of 12BE. The formation of a significant fraction of circular oligomers in PEG/EDC/6SE mixtures is demonstrated

by the combination of two observations. NMR ^{13}C signals from the anomeric carbons indicate that virtually all 6SE 5′ terminal phosphates are involved in phosphodiester bonds (see Supporting Information). This evidence demonstrates that the ligation yield is much larger than the one determined by product length distribution, a condition which is found if a significant fraction of the oligomers is circular. Indeed, PAGE analysis reveals a number of bands larger than the expected multiples of 6SE (Figure 2b, green frame). We thus performed enzymatic degradation of the reaction products of 6SE by SVP (Snake Venom Phosphodiesterase I), an enzyme known to degrade nucleic acid molecules with an efficiency larger for linear than for circular strands.[62−64] We dispersed 0.1 mU of SVP in the reaction products of PEG/EDC/6SE with c_{PEG} = 250 mg/mL and compared the effects of enzymatic deg-radation at different times, as shown in Figure 5. Quite clearly, the SVP action on the various products is uneven, as apparent in the section zoomed in panel b and analyzed in panel c. Two of the bands (red squares and green triangles) are almost completely degraded after 5 min, while the intensity of the third (blue dots) decreases in the same time of only about 20%. This result enables the identification of the former as linear and the latter as circular products, thus justifying the large number of bands. It also indicates that they alternate in the gel, which is consistent with the attribution resulting from the 2D PAGE gels in 12BE reaction products (see Supporting Information): linear strands have larger electrophoretic mobility than circular chains of equal mass.

All this evidence supports the notion that the molecular template provided by LC ordering is a crucial factor in guiding polymerization toward long polymers, since in its absence intramolecular cyclization is the dominant reaction pathway.

CONCLUSION

In this work, we have demonstrated the templating effect of LC ordering in solutions of RNA oligomers. Our results provide an example of self-synthesizing material, in which self-assembly

promotes the formation of products that in turn stabilize the assembly. At the same time, our results strengthen the notion that LC ordering may have had a role in favoring the first emergence of RNA polymers as a product of ligation between short oligomers.

Our study has been organized in steps.

(1) We have found that solutions of RNA oligomers develop LC ordering in a wide range of chain lengths, concentrations, and temperatures. We thoroughly investigated the phase diagrams, identifying the conditions more amenable to a ligation reaction. Specifically, we found that RNA oligomers can be concentrated and segregated using PEG as a molecular crowding agent and that the presence of Mg^{2+} ions has a stabilizing effect on RNA LC phases, significantly promoting phase separation in RNA-PEG mixtures.

(2) We thus studied the ligation reaction in mixtures in which RNA oligomers are organized in fluid microdomains, either liquid or LC, coexisting with a liquid phase rich in PEG that acts as a reservoir for EDC, a carbodiimide agent that activates the terminal phosphates, and as a sink for waste. We have found that, as LC ordering appears, the length of the reaction products markedly increases, to the point that almost 50% of the resulting chains is more than five-times longer than the starting oligomers. The mean average length of the products increases from 24 bases (twice the initial 12mer) to 72 bases (6-times) when the reaction is carried in isotropic and LC phase, respectively.

(3) We have thus focused on understanding which, among the various steps involved in the reaction, is the one that promotes the enhanced ligation and why. By combining the study of the reaction kinetics in various conditions, of the partitioning of the solutes in the coexisting phases, of the viscosity of the RNA solutions, we managed to disentangle the various contributions in the ligation reactions and show that the appearance of LC ordering markedly enhances the last step of the reaction, that is, the formation of phosphodiester bonds between the activated 5′ terminal phosphate and the 3′ (or 2′) OH terminus. This result is a proof that the effect of LC ordering is indeed a consequence of the template provided by the collective molecular ordering. The geometrical constraint exerted on each duplex lowers the entropic cost of covalent bonding, thus speeding up the reaction. Our observations suggest an analogy between LC templating and enzyme-promoted chemical ligation between the termini of nucleic acid chains. Indeed, the action of ligases relies on the combination of various factors, including (i) the use of molecular energy sources, such as ATP, to produce a reactive intermediate (ii) the positional constraint on the two termini following their binding to the ligase, (iii) the reduction of the free energy of the transition state provided by the loss of entropy of the reactants and thanks to the electrochemical environment in the enzyme active site. The LC-promoted abiotic ligation reaction pathway bears relevant analogies to the working scheme of ligase enzymes: (i) the energy source is here provided by EDC, whose reaction produces a reactive intermediate (ii) kept in close proximity to a reactive terminus by the LC milieu, (iii) thus reducing the entropy of activation of

the transition state in the reaction step leading to the formation of a new phosphodiester bond.

(4) We have found that LC ordering strongly disfavors the formation of cyclic products, which instead appear to be the favored outcome when reaction between duplex termini takes place in a liquid phase. Indeed, the formation of circular products is a potential dead-end of abiotic polymerization and thus a fatal obstacle in the spontaneous emergence of long RNA polymers on the early Earth. This is a recognized problem for the spontaneous polymerization of unstructured fluids. Previous studies have proposed the presence of interbase intercalating molecules as a possible mechanism disfavoring cyclization.[16]

Our results indicate that both templating effect and polymerization in linear chains are promoted by supramolecular ordering with no need of external guiding constraints such as intercalators, lipids or mineral surfaces. The efficacy of LC ordering in promoting the formation of linear RNA polymers is the result of two combined factors: molecular order and fluidity. The onset of uniaxial LC ordering brings about a symmetry breaking that is reflected at the microscopic level as a molecular field stabilizing linear aggregates and suppressing supramolecular clusters having alternative geometries such as circular or disordered aggregates. Fluidity enables the diffusion, albeit slowed down by the increased viscosity, of reactant and waste molecules and the circulation of the duplexes.

To which extent the conditions here considered are compatible with the primordial Earth is of course a matter of speculation. The presence of short oligomers could have been promoted by lipid assisted polymerization of single nucleotides.[37,39,40] The concentration necessary for LC templating could have been achieved either in inland hydrothermal water ponds fed by hot springs, by hydration−dehydration cycles,[12,49] or in oceanic hydrothermal pores by accumulation of RNA molecules driven by thermophoresis.[50,65] Both sites are characterized by temperature variations compatible with those explored in this work.[49,50] Ionic strength and Mg^{2+} concentration used in our experiments are also in line with the expected composition of prebiotic water. In modern oceans, the Mg^{2+} concentration is about 54 mM, while in the anoxic water of primordial oceans it is estimated to be around 10 mM[66,67] and in the sweet water of inland thermal springs is of the order of few millimolar.[49,66] PEG and EDC are instead not prebiotically plausible molecules. Indeed, we have used them not because they represent a realistic model, but for their effectiveness in generating conditions that could have been produced in many other ways. PEG has been used in this work as a molecular crowder. In general, crowding can result from dehydration or from phase separation. PEG induces and enables easy control of aqueous two-phase systems, which have been considered as model systems for coacervates, which are thought to have played a crucial role in selecting and clustering molecules in prebiotic environments.[68−71] Other molecular species, such as polysaccharides, cationic oligopeptides, and low molecular weight polycations (as spermine and spermidine), have also been found to induce coacervation,[72,73] providing a set of more prebiotically plausible alternatives to PEG. Analogously, while EDC has been chosen for its fast reactivity to enable demonstrating the intimate connection between LC ordering and linear polymerization, other phosphate activating molecules, such as cyanogen bromide[74] and

N-cyanoimidazole,[75] could represent prebiotically plausible alternatives to EDC.

Overall, our experiments provide a proof of concept of LC templating: indeed, should there have been an "RNA World" on the primordial Earth, our results suggest that it could have been preceded by a "LC World", whose appearance requires much shorter chains, in which supramolecular ordering provided molecular selection and geometrical template for the polymerization of long linear RNA chains.

METHODS

Solid Phase RNA Synthesis and Purification. RNA-oligonucleotides were assembled on an Äkta10 Oligopilot synthesizer (Amersham Biosciences; GE Healthcare, Freiburg, D) in a 1.2 mL fixed volume column using standard RNA phosphoramidites. Phosphoramidites were purchased from Proligo (Hambug,Germany, D-RNAphosphoramidites) and ChemGenes (Wilmington,USA, L-RNA phosphoramidites). Base-loaded solid support was purchased from Prime Synthesis (Aston, PA, USA). Synthesis was started from base-loaded CPG, pore size 600 Å. For coupling (12 min per cycle), 0.6 M ethylthiotetrazole (Azide Chemical Co., Ltd., Anzhen, Wuxi, CN) in acetonitrile, and 6 equiv of the respective 0.2 M phosphoramidite solution in acetonitrile were used.

A capping-oxidation-capping cycle was applied. Standard solvents and reagents for oligonucleotide synthesis were purchased from Biosolve (Valkenswaard, NL), Proligo (Hamburg, D), VWR (Karlsruhe, D), or Sigma-Aldrich (Taufkirchen, D). Cleavage and deprotection were achieved according to Wincott et al.[76] with minor alterations. In detail, upon completion of the automated synthesis, the CPG-bound oligonucleotide (10–15 μmol) was briefly dried and transferred into a glass bottle. Twenty milliliters of aq. MeNH$_2$ (40%) were added, and the suspension was gently agitated at room temperature. After 90 min, the slurry was filtered, and the residual CPG washed several times with aq. EtOH (50%). The combined filtrates were concentrated and lyophilized. For the removal of the 2′ TBDMS groups, the dry crude product was dissolved in 1.5 mL of DMSO followed by 0.75 mL of NEt$_3$ and 1.0 mL of NEt$_3$·3HF. This mixture was gently agitated at 65 °C for 2 h. After cooling to room temperature, 20 mL of n-BuOH was added and the resulting precipitate was collected and washed with acetone. Crude oligonucleotides were purified by ion exchange HPLC (IEX) followed by Salt removal size exclusion chromatography (NAP10, GE-Healthcare, Freiburg, Germany). The final products were dried by lyophilization and stored at −20 °C until further use. Purity and identity were confirmed by IEX- and RP-HPLC and LC−MS.

Microscope Cells Preparation. Lyophilized 5′-phosphate RNA was resuspended in Milli-Q water at a concentration of 50 mg/mL and stored at −20 °C to slow down spontaneous hydrolysis. Samples for microscope observations (microscope cells) were prepared placing one or multiple drops (0.5−1 μL) of the RNA solution on a glass slide, waiting for them to dry and resuspending them in an appropriate volume of solutions based on Milli-Q water, MgCl$_2$ and buffered PEG8k to reach the final desired concentration directly on the slide.

To observe these sample in standard bright field, polarized light and fluorescence microscopy, a second glass slide was placed on the top of the solution and kept at a controlled distance using 10−20 μm silica spacer rods. The LC domains explored in this study are thus much larger than both molecular and aggregate sizes. The cells were sealed with fluorinated oil to keep the sample concentration stable over long periods of time at high temperature.

C$_{RNA}$ versus Temperature Phase Diagram Measurement. The measurement of RNA concentration in microscope cells was performed through microscope-based interferometry.[21] One or multiple drops of concentrated RNA (initial concentration = 50 mg/mL) was placed on a high-refractive index glass (n = 1.62), sandwiched, and slowly concentrated through evaporation at room temperature before being sealed with fluorinated oil at different times to produce an array of different final concentrations. The liquid crystalline solutions

prepared in this way were equilibrated through multiple thermal cycles, taking note of the different phases exhibited as a function of temperature, and finally brought to isotropic phase, where the spectrum of the light reflected in the RNA solution was acquired. The multiple reflection fringes due to the interference from the parallel glass plates were systematically measured in various spots within the isotropic RNA solution and in the fluorinated oil (of known refractive index) present at the edges of the cell. This set of data enabled determining the exact thickness of the cell and the refractive index of the RNA solution, and thus the RNA concentration c$_{RNA}$ through a calibration curve (n = n$_0$ + dn/dc × c), with n$_0$ the refractive index of the solution at zero RNA concentration and dn/dc = 0.136 cm^3/g).

1D PAGE. Two micrograms of reaction product in 5 μL was mixed with 3 μL of loading solution (30% glycerol) and 10 μL of formamide and loaded in a 7 M urea denaturing 15% polyacrylamide gel (size 20 × 20 × 0.1 cm^3). Gels were run in a Protean apparatus (Biorad) in TBE buffer at 250 V for approximately 2.5 h. SYBR Gold 1x in TBE buffer was used as staining for 5 min and gel images were acquired in .tif format using a Gel Doc EZ Imager (Biorad).

2D PAGE. A single lane was excised from the 1D gel and stacked on top of a second 7 M urea denaturing 20% polyacrylamide gel (size 20 × 20 × 0.1 cm^3). The two gels were sealed together using fresh 20% polyacrylamide gel. The second dimension was run at 300 V for approximately 4 h. Staining and image acquisition were done as previously described.

Analysis of Polymerization Yield. Lane profiles of the acquired gel images were estimated using MacBiophotonics ImageJ software. Assuming linearity between the fluorescence intensity of the peaks and the RNA concentration, it is possible to measure the product in each gel band. The cumulative fluorescence signal of the different length product is fitted with Flory Model equation for a simple polymerization[56] h(p,x) = 1 − p^x − x × p^x + x × p$^{(1+x)}$, describing the products length distribution of a system where monomers have the same probability p to bound and p is independent from polymer length. See Supporting Information for a more detailed description. Each experimental condition has been repeated three times in distinct microscope cells and the measured yields have been found to be consistent.

pH Determination. The pH of RNA/PEG8k/MgCl$_2$/EDC mixtures was measured using an Orion 9810BN Micro pH Electrode. The pH determined in this way is equal to 7.3. Unless otherwise stated all the spectrophotometric measurements have been performed using an excess of HEPES pH 7.3 buffer to match the mixtures pH.

Determination of Concentrations in PEG-Separated Phases. To evaluate the concentration of RNA in the PEG-rich phase, we measured the relative volumes filled by the RNA-rich phase as a function of c$_{RNA}$ (T = 25 °C, 60 mM MgCl$_2$) in c$_{PEG}$ = 150 mg/mL and c$_{PEG}$ = 200 mg/mL mixtures. As c$_{RNA}$ concentration is increased, the volume fraction occupied by birefringent domains increases linearly, and the RNA concentration in the domains can be obtained from the proportionality coefficient. See Supporting Information for a more detailed description.

FRAP-Based Measurement of Samples Viscosity. We used sodium fluorescein as fluorophore to probe the viscosity of the RNA and RNA-PEG mixtures due to its hydrodynamic radius (r = 0.45 nm) being comparable to the one of EDC. Bleaching was induced focusing a Xenon Lamp source over an area of 80 μm−400 μm of radius on the sample with a high magnification objective (20× or 50×). Recovery was recorded using a low magnification objective (4× or 10×) and neutral filters to reduce bleaching during the acquisition. Data analysis was performed according to Soumpasis model.[77] See Supporting Information for more details.

Determination of k_h and k_1 Using UV-Spectrophotometry. k_h can be readily determined by monitoring the time evolution of the EDC absorbance spectrum and the corresponding appearance and growth of the absorption spectrum of EDU as shown in Figure S11, left-hand panel. Normalized profiles were fitted using a simple exponential decay e$^{−kt}$ (right-hand panels). We obtain $k_h \approx 0.033$ h$^{−1}$. To measure k_1, we monitored the decay of EDC in mixtures with 5′phosphate ribose (P for simplicity), adding various amounts of P and fitting the decay of the amplitude of the EDC spectrum with

$c \times e^{(t(-[P] \cdot k1 - kh))}$, where c is the initial EDC concentration, $[P]$ is the concentration of 5′phosphate ribose, and k_h is held as previously determined. See Supporting Information for more details.

Modeling the Ligation Reaction. The reaction described in Figure 4 of the main text can be modeled by a set of coupled linear reactions (see SI eq S11). These equations, combined with the conservation constraints between species, can be solved to obtain the expected evolution of all species. Since no analytical solution is available, we solved them numerically using ODE45 function in MATLAB, searching for the best k_2 to k_L ratio ($R = k_2 / k_L$) to fit our experimental data (implemented via the *fminsolve* function) for product formation obtained from PAGE gels as described above. The value of the parameter R that best reproduces our experimental data in the mixed state is $R \approx 1.11 \times 10^{-2}$. See Supporting Information for more details. Eq S11 is conceived for a homogeneous solution. When the system phase separates, the reaction takes place simultaneously in the two phases. The set of coupled equations can be modified to take into account this different condition introducing the parameters F_{in} and F_{out} as the volume fractions of the two phases (see SI eq S11) and incorporating into the system the larger viscosity in the RNA-rich phases. See Supporting Information for more details.

NMR Characterization of Ligation Products. We performed high resolution nuclear magnetic resonance (NMR) spectroscopy on two 6SE samples: a 6SE stock solution (6SE) and a 6SE sample that underwent 48 h ligation reaction in LC phase (6SE-LIG), c_{PEG} = 200 mg/mL and EDC 30X. ^{1}H NMR analyses were performed at 500 MHz with a Bruker FT-NMR AVANCE DRX500 spectrometer using a 5 mm z-PFG (pulsed field gradient) broadband reverse probe at 298 and 343 K, and ^{13}C NMR spectra were collected at 125.76 MHz. The signals were unambiguously assigned by 2D COSY, NOESY (only 298 K), and HSQC experiments.[78,79] ^{1}H NMR chemical shifts are reported as δ (ppm) relative to residual HDO fixed at 4.705 and 4.716 ppm for spectra recorded at 298 and 343 K, respectively. See Supporting Information for more details.

Enzymatic Digestion of 6SE. The 6SE ligation with EDC was treated with SVP, Snake Venom Phosphodiesterase I from *Crotalus adamanteus* (Sigma-Aldrich). The lyophilized enzyme was resuspended in 4 mL of reaction buffer (Tris-HCl 10 mM pH 8.8, NaCl 25 mM, MgCl$_2$ 10 mM) and 1 μL was added to a 9 μL solution of the ligated RNA diluted in reaction buffer. Final RNA concentration in digestion reaction is ~0.1−0.2 mg/mL, final enzyme activity is ~0.01 U/mL. The reaction was run at room temperature and stopped after 5′, 10′, 30′, and 60′ adding 2 μL of 55 mM EDTA pH 8 to the mixtures and immediately storing them at −20 °C. See Supporting Information for more details.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsnano.8b05821.

Additional figures, detailed description of 2D PAGE gels, enzymatic digestion of 6SE, analysis of polymerization yield, characterization of reaction mixtures, determination of reaction rates, NMR characterization of ligation products (PDF)

Full NMR spectra (PDF)

AUTHOR INFORMATION

Corresponding Author
*E-mail: tommaso.bellini@unimi.it.

ORCID Ⓞ
Tommaso Bellini: 0000-0003-4898-4400

Author Contributions
M.T., T.P.F., G.P.S., G.Z., N.A.C., and T.B. conceived the experiments; L.B. and S.K. synthesized and purified the oligomers; M.T., T.P.F., and A.C. measured phase diagrams and defined the ligation conditions; M.T. and T.P.F. performed ligation experiments and enzymatic digestion; M.T., E.M.P., and R.A. performed PAGE; M.T. performed UV absorbance and FRAP experiments; M.T. and T.B. analyzed the data; T.P.F. and D.C. performed the NMR measurements; M.T., T.P.F., G.Z., N.A.C., and T.B. wrote the manuscript.

REFERENCES
(1) Bissette, A. J.; Fletcher, S. P. Mechanisms of Autocatalysis. *Angew. Chem., Int. Ed.* **2013**, *52*, 12800−12826.

(2) Li, J.; Carnall, J. M. A.; Stuart, M. C. A.; Otto, S. Hydrogel Formation upon Photoinduced Covalent Capture of Macrocycle Stacks from Dynamic Combinatorial Libraries. *Angew. Chem., Int. Ed.* **2011**, *50*, 8384−8386.

(3) Omosun, T. O.; Hsieh, M. C.; Childers, W. S.; Das, D.; Mehta, A. K.; Anthony, N. R.; Pan, T.; Grover, M. A.; Berland, K. M.; Lynn, D. G. Catalytic Diversity in Self-Propagating Peptide Assemblies. *Nat. Chem.* **2017**, *9*, 805−809.

(4) Makhlynets, O. V.; Gosavi, P. M.; Korendovych, I. V. Short Self-Assembling Peptides Are Able to Bind to Copper and Activate Oxygen. *Angew. Chem., Int. Ed.* **2016**, *55*, 9017−9020.

(5) Rubinov, B.; Wagner, N.; Matmor, M.; Regev, O.; Ashkenasy, N.; Ashkenasy, G. Transient Fibril Structures Facilitating Non-enzymatic Self-Replication. *ACS Nano* **2012**, *6*, 7893−7901.

(6) Rout, S. K.; Friedmann, M. P.; Riek, R.; Greenwald, J. A Prebiotic Template-Directed Peptide Synthesis Based on Amyloids. *Nat. Commun.* **2018**, *9*, 234.

(7) Ashkenasy, G.; Hermans, T. M.; Otto, S.; Taylor, A. F. Systems Chemistry. *Chem. Soc. Rev.* **2017**, *46*, 2543−2554.

(8) Budin, I.; Szostak, J. W. Expanding Roles for Diverse Physical Phenomena During the Origin of Life. *Annu. Rev. Biophys.* **2010**, *39*, 245−263.

(9) Cafferty, B. J.; Gállego, I.; Chen, M. C.; Farley, K. I.; Eritja, R.; Hud, N. V. Efficient Self-Assembly in Water of Long Noncovalent Polymers by Nucleobase Analogues. *J. Am. Chem. Soc.* **2013**, *135*, 2447−2450.

(10) Davis, J. G.; Gierszal, K. P.; Wang, P.; Ben-Amotz, D. Water Structural Transformation at Molecular Hydrophobic Interfaces. *Nature* **2012**, *491*, 582−585.

(11) Szostak, J. W. The Narrow Road to the Deep Past: In Search of the Chemistry of the Origin of Life. *Angew. Chem., Int. Ed.* **2017**, *56*, 11037−11043.

(12) Pearce, B. K. D.; Pudritz, R. E.; Semenov, D. A.; Henning, T. K. Origin of the RNA World: The Fate of Nucleobases in Warm Little Ponds. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 11327−11332.

(13) Powner, M. W.; Gerland, B.; Sutherland, J. D. Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions. *Nature* **2009**, *459*, 239−242.

(14) Mamajanov, I.; Macdonald, P. J.; Ying, J.; Duncanson, D. M.; Dowdy, G. R.; Walker, C. A.; Engelhart, A. E.; Fernández, F. M.;

Grover, M. A.; Hud, N. V.; Schork, F. J. Ester Formation and Hydrolysis during Wet-Dry Cycles: Generation of Far-from-Equilibrium Polymers in a Model Prebiotic Reaction. *Macromolecules* **2014**, *47*, 1334−1343.

(15) Toppozini, L.; Dies, H.; Deamer, D. W.; Rheinstädter, M. C. Adenosine Monophosphate Forms Ordered Arrays in Multilamellar Lipid Matrices: Insights into Assembly of Nucleic Acid for Primitive Life. *PLoS One* **2013**, *8*, 1−8.

(16) Horowitz, E. D.; Engelhart, A. E.; Chen, M. C.; Quarles, K. a.; Smith, M. W.; Lynn, D. G.; Hud, N. V. Intercalation as a Means to Suppress Cyclization and Promote Polymerization of Base-Pairing Oligonucleotides in a Prebiotic World. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 5288−5293.

(17) Kuruvilla, E.; Schuster, G. B.; Hud, N. V. Enhanced Nonenzymatic Ligation of Homopurine Miniduplexes: Support for Greater Base Stacking in a Pre-RNA World. *ChemBioChem* **2013**, *14*, 45−48.

(18) Benner, S. A.; Kim, H. J.; Carrigan, M. A. Asphalt, Water, and the Prebiotic Synthesis of Ribose, Ribonucleosides, and RNA. *Acc. Chem. Res.* **2012**, *45*, 2025−2034.

(19) Fraccia, T. P.; Smith, G. P.; Zanchetta, G.; Paraboschi, E.; Yi, Y.; Walba, D. M.; Dieci, G.; Clark, N. A.; Bellini, T. Abiotic Ligation of DNA Oligomers Templated by Their Liquid Crystal Ordering. *Nat. Commun.* **2015**, *6*, 6424.

(20) Jones, M. R.; Seeman, N. C.; Mirkin, C. A. Programmable Materials and the Nature of the DNA Bond. *Science* **2015**, *347*, 1260901−1260901.

(21) Nakata, M.; Zanchetta, G.; Chapman, B. D.; Jones, C. D.; Cross, J. O.; Pindak, R.; Bellini, T.; Clark, N. A. End-to-End Stacking and Liquid Crystal Condensation of 6 to 20 Base Pair DNA Duplexes. *Science* **2007**, *318*, 1276−1279.

(22) Zanchetta, G.; Nakata, M.; Buscaglia, M.; Clark, N. A.; Bellini, T. Liquid Crystal Ordering of DNA and RNA Oligomers with Partially Overlapping Sequences. *J. Phys.: Condens. Matter* **2008**, *20*, 494214−494219.

(23) Bellini, T.; Zanchetta, G.; Fraccia, T. P.; Cerbino, R.; Tsai, E.; Smith, G. P.; et al. Liquid Crystal Self-Assembly of Random-Sequence DNA Oligomers. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1−6.

(24) Fraccia, T. P.; Smith, G. P.; Bethge, L.; Zanchetta, G.; Nava, G.; Klussmann, S.; Clark, N. A.; Bellini, T. Liquid Crystal Ordering and Isotropic Gelation in Solutions of Four-Base-Long DNA Oligomers. *ACS Nano* **2016**, *10*, 8508−8516.

(25) Smith, G. P.; Fraccia, T. P.; Todisco, M.; Zanchetta, G.; Zhu, C.; Hayden, E.; Bellini, T.; Clark, N. A. Backbone-Free Duplex-Stacked Monomer Nucleic Acids Exhibiting Watson − Crick Selectivity. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E7658−E7664.

(26) Fraccia, T. P.; Smith, G. P.; Clark, N. A.; Bellini, T. Liquid Crystal Ordering of Four-Base-Long DNA Oligomers with Both G−C and A−T Pairing. *Crystals* **2018**, *8*, 494214−494219.

(27) Zanchetta, G.; Nakata, M.; Buscaglia, M.; Bellini, T.; Clark, N. A. Phase Separation and Liquid Crystallization of Complementary Sequences in Mixtures of NanoDNA Oligomers. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 1111−1117.

(28) Gilbert, W. The RNA World. *Nature* **1986**, *319*, 618.

(29) Joyce, G. F. The Antiquity of RNA-Based Evolution. *Nature* **2002**, *418*, 214−221.

(30) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* **1982**, *31*, 147−157.

(31) Scott, W. G.; Finch, J. T.; Klug, A. The Crystal Structure of an AII-RNA hammerhead Ribozyme: A Proposed Mechanism for RNA Catalytic Cleavage. *Cell* **1995**, *81*, 991−1002.

(32) Johnston, W. K.; Unrau, P. J.; Lawrence, M. S.; Glasner, M. E.; Bartel, D. P. RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science* **2001**, *292*, 1319−1325.

(33) Robertson, M. P.; Joyce, G. F. The Origins of the RNA World. *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003608.

(34) De Duve, C. *Singularities - Landmarks on the Pathways of Life*; Cambridge University Press, 2005.

(35) Miyakawa, S.; Ferris, J. P. Sequence- and Regioselectivity in the Montmorillonite-Catalyzed Synthesis of RNA. *J. Am. Chem. Soc.* **2003**, *125*, 8202−8208.

(36) Huang, W.; Ferris, J. P. One-Step, Regioselective Synthesis of up to 50-Mers of RNA Oligomers by Montmorillonite Catalysis. *J. Am. Chem. Soc.* **2006**, *128*, 8914−8919.

(37) Rajamani, S.; Vlassov, A.; Benner, S.; Coombs, A.; Olasagasti, F.; Deamer, D. Lipid-Assisted Synthesis of RNA-like Polymers from Mononucleotides. *Origins Life Evol. Biospheres* **2008**, *38*, 57−74.

(38) Mansy, S. S.; Schrum, J. P.; Krishnamurthy, M.; Tobé, S.; Treco, D. a.; Szostak, J. W. Template-Directed Synthesis of a Genetic Polymer in a Model Protocell. *Nature* **2008**, *454*, 122−125.

(39) Deguzman, V.; Vercoutere, W.; Shenasa, H.; Deamer, D. Generation of Oligonucleotides Under Hydrothermal Conditions by Non-Enzymatic Polymerization. *J. Mol. Evol.* **2014**, *78*, 251−262.

(40) Mungi, C. V.; Rajamani, S. Characterization of RNA-Like Oligomers from Lipid-Assisted Nonenzymatic Synthesis: Implications for Origin of Informational Molecules on Early Earth. *Life* **2015**, *5*, 65−84.

(41) Von Kiedrowski, G. A Self-Replicating Hexadeoxynucleotide. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 932−935.

(42) Zielinski, W. S.; Orgel, L. E. Autocatalytic Synthesis of a Tetranucleotide Analogue. *Nature* **1987**, *327*, 346−347.

(43) Inoue, T.; Orgel, L. E. Substituent Control of the Poly (C)-Directed Oligomerization of Guanosine 5′-Phosphoroimidazolid. *J. Am. Chem. Soc.* **1981**, *103*, 7666−7667.

(44) Schrum, J. P.; Ricardo, A.; Krishnamurthy, M.; Blain, J. C.; Szostak, J. W. Efficient and Rapid Template-Directed Nucleic Acid Copying Using 2′-Amino-2′,3′-Dideoxyribonucleoside-5′- Phosphorimidazolide Monomers. *J. Am. Chem. Soc.* **2009**, *131*, 14560−14570.

(45) Taran, O.; Thoennessen, O.; Achilles, K.; Von Kiedrowski, G. Synthesis of Information-Carrying Polymers of Mixed Sequences from Double Stranded Short Deoxynucleotides. *J. Syst. Chem.* **2010**, *1*, 9−24.

(46) Fraccia, T. P.; Zanchetta, G.; Rimoldi, V.; Clark, N. a.; Bellini, T. Evidence of Liquid Crystal−Assisted Abiotic Ligation of Nucleic Acids. *Origins Life Evol. Biospheres* **2015**, *45*, 51−68.

(47) Calladine, C. R.; Drew, H. R.; Luisi, B. F.; Travers, A. A.; Calladine, C. R.; Drew, H. R.; Luisi, B. F.; Travers, A. A. Chapter 3 − Different Kinds of Double Helix. In *Understanding DNA*; Elsevier, 2004; pp 39−63.

(48) Zanchetta, G.; Bellini, T.; Nakata, M.; Clark, N. A. Physical Polymerization and Liquid Crystallization of RNA Oligomers. *J. Am. Chem. Soc.* **2008**, *130*, 12864−12865.

(49) Damer, B.; Deamer, D. Coupled Phases and Combinatorial Selection in Fluctuating Hydrothermal Pools: A Scenario to Guide Experimental Approaches to the Origin of Cellular Life. *Life* **2015**, *5*, 872−887.

(50) Baaske, P.; Weinert, F. M.; Duhr, S.; Lemke, K. H.; Russell, M. J.; Braun, D. Extreme Accumulation of Nucleotides in Simulated Hydrothermal Pore Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9346−9351.

(51) Rossi, M.; Zanchetta, G.; Klussmann, S.; Clark, N. a.; Bellini, T. Propagation of Chirality in Mixtures of Natural and Enantiomeric Dna Oligomers. *Phys. Rev. Lett.* **2013**, *110*, 107801.

(52) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. Nonenzymatic, Template-Directed Ligation of Oligoribonucleotides Is Highly Regioselective for the Formation of 3 ′ - 5 ′ Phosphodiester Bonds. *J. Am. Chem. Soc.* **1996**, *118*, 3340−3344.

(53) Sawai, H.; Wada, M.; Kouda, T.; Ozaki, A. N. Nonenzymatic Ligation of Short-Chained 2′-5′- or 3′-5′-Linked Oligoribonucleotides on 2′-5′- or 3′-5′-Linked Complementary Templates. *ChemBioChem* **2006**, *7*, 605−611.

(54) Lohrmann, R.; Orgel, L. E. Preferential Formation of (2′-5′)-Linked Internucleotide Bonds in Non-Enzymatic Reactions. *Tetrahedron* **1978**, *34*, 853−855.

(55) Sokolova, N. I.; Ashirbekova, D. T.; Dolinnaya, N. G.; Shabarova, Z. A. Chemical Reactions within DNA Duplexes. Cyanogen Bromide as an Effective Oligodeoxyribonucleotide Coupling Agent. *FEBS Lett.* **1988**, *232*, 153−155.

(56) Flory, J. Molecular Size Distribution in Linear Condensation Polymers. *J. Am. Chem. Soc.* **1936**, *58*, 1877−1885.

(57) Voit, B.; Haag, R.; Appelhans, D.; Welzel, P. B. Chapter 4.1.1 − Structural Characterization. In *Bio- and Multifunctional Polymer Architectures: Principles, Synthetic Methods, and Applications*; John Wiley & Sons, 2016; p 116.

(58) Grabarek, Z.; Gergely, J. Zero-Length Crosslinking Procedure with the Use of Active Esters ☆. *Anal. Biochem.* **1990**, *185*, 131−135.

(59) Nowacka, M.; Jackowiak, P.; Rybarczyk, A.; Magacz, T.; Strozycki, P. M.; Barciszewski, J.; Figlerowicz, M. 2D-PAGE as an Effective Method of RNA Degradome Analysis. *Mol. Biol. Rep.* **2012**, *39*, 139−146.

(60) Fradin, A.; Gruhl, H.; Feldmann, H. Mapping of Yeast TRNAs by Two-Dimensional Electrophoresis on Polyacrylamide Gels. *FEBS Lett.* **1975**, *50*, 185−189.

(61) Hassall, J.; MacDonald, P.; Cordero, T.; Rostain, W.; Jaramillo, A. Design and Characterization of Topological Small RNAs. In *RNA Scaffolds. Methods in Molecular Biology*; Springer, 2015; pp 149−167.

(62) Zaphiropoulos, P. G. Circular RNAs from Transcripts of the Rat Cytochrome P450 2C24 Gene: Correlation with Exon Skipping. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 6536−6541.

(63) Gao, H.; Yang, M.; Patel, R.; Cook, A. F. Circularization of Oligonucleotides by Disulfide Bridge Formation. *Nucleic Acids Res.* **1995**, *23*, 2025−2029.

(64) Kowalski, D. A Procedure for the Quantitation of Relaxed Closed Circular DNA in the Presence of Superhelical DNA: An Improved Fluorometric Assay for Nicking-Closing Enzyme. *Anal. Biochem.* **1979**, *93*, 346−354.

(65) Mast, C. B.; Schink, S.; Gerland, U.; Braun, D. Escalation of Polymerization in a Thermal Gradient. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 8030−8035.

(66) Mulkidjanian, A. Y.; Bychkov, A. Y.; Dibrova, D. V.; Galperin, M. Y.; Koonin, E. V. Origin of First Cells at Terrestrial, Anoxic Geothermal Fields. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E821−E830.

(67) Holm, N. G. The Significance of Mg in Prebiotic Geochemistry. *Geobiology* **2012**, *10*, 269−279.

(68) Jia, T. Z.; Hentrich, C.; Szostak, J. W. Rapid RNA Exchange in Aqueous Two-Phase System and Coacervate Droplets. *Origins Life Evol. Biospheres* **2014**, *44*, 1−12.

(69) Keating, C. D. Aqueous Phase Separation as a Possible Route to Compartmentalization of Biological Molecules. *Acc. Chem. Res.* **2012**, *45*, 2114−2124.

(70) Strulson, C. A.; Molden, R. C.; Keating, C. D.; Bevilacqua, P. C. RNA Catalysis through Compartmentalization. *Nat. Chem.* **2012**, *4*, 941.

(71) Lazcano, A. Historical Development of Origins Research. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, a002089.

(72) Poudyal, R. R.; Pir Cakmak, F.; Keating, C. D.; Bevilacqua, P. C. Physical Principles and Extant Biology Reveal Roles for RNA-Containing Membraneless Compartments in Origins of Life Chemistry. *Biochemistry* **2018**, *57*, 2509−2519.

(73) Matthew, H. W.; Salley, S. O.; Peterson, W. D.; Klein, M. D. Complex Coacervate Microcapsules for Mammalian Cell Culture and Artificial Organ Development. *Biotechnol. Prog.* **1993**, *9*, 510−519.

(74) Dolinnaya, N. G.; Sokolova, N. I.; Ashirbekova, D. T.; Shabarova, Z. a. The Use of BrCN for Assembling Modified DNA Duplexes and DNA-RNA Hybrids; Comparison with Water-Soluble Carbodiimide. *Nucleic Acids Res.* **1991**, *19*, 3067−3072.

(75) Mariani, A.; Sutherland, J. D. Non-Enzymatic RNA Backbone Proofreading through Energy-Dissipative Recycling. *Angew. Chem., Int. Ed.* **2017**, *56*, 6563−6566.

(76) Wincott, F.; Direnzo, A.; Shaffer, C.; Grimm, S.; Tracz, D.; Workman, C.; Sweedler, D.; Gonzalez, C.; Scaringe, S.; Usman, N. Synthesis, Deprotection, Analysis and Purification of RNA and Ribosomes. *Nucleic Acids Res.* **1995**, *23*, 2677−2684.

(77) Soumpasis, D. M. Theoretical Analysis of Fluorescence Photobleaching Recovery Experiments. *Biophys. J.* **1983**, *41*, 95−97.

(78) Parella, T. Pulse Program Catalogue : Experiments. *Pulse Progr. Cat* **2006**, 1−249.

(79) Kessler, H.; Gehrke, M.; Griesinger, C. Two-Dimensional NMR Spectroscopy: Background and Overview of the Experiments. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 490−536.

CHAPTER C

Paper 2

Backbone–free duplex-stacked monomer nucleic acids exhibiting Watson–Crick selectivity.

I have contributed to this work as follows:

- Experimental design;

- Phase diagrams characterization through polarized optical microscopy.

Backbone-free duplex-stacked monomer nucleic acids exhibiting Watson–Crick selectivity

Gregory P. Smith[a,1], Tommaso P. Fraccia[b,c,1], Marco Todisco[b], Giuliano Zanchetta[b], Chenhui Zhu[d], Emily Hayden[a], Tommaso Bellini[b,2], and Noel A. Clark[a,2]

[a]Department of Physics and Soft Materials Research Center, University of Colorado Boulder, Boulder, CO 80309-0390; [b]Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, I-20090 Segrate, Italy; [c]Dipartimento di Promozione delle Scienze Umane e della Qualità della Vita, Università Telematica San Raffaele, I-00166 Roma, Italy; and [d]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

We demonstrate that nucleic acid (NA) mononucleotide triphosphates (dNTPs and rNTPs), at sufficiently high concentration and low temperature in aqueous solution, can exhibit a phase transition in which chromonic columnar liquid crystal ordering spontaneously appears. Remarkably, this polymer-free state exhibits, in a self-assembly of NA monomers, the key structural elements of biological nucleic acids, including: long-ranged duplex stacking of base pairs, complementarity-dependent partitioning of molecules, and Watson–Crick selectivity, such that, among all solutions of adenosine, cytosine, guanine, and thymine NTPs and their binary mixtures, duplex columnar ordering is most stable in the A-T and C-G combinations.

nucleic acid | nucleoside triphosphate | liquid crystal | double helix | prebiotic evolution

The transfer and memory of genetic information in life is based on the supramolecular structuring of nucleic acid (NA) polymers in solution (1). Selective interactions between the side groups (bases) on pairs of NA chains direct molecular self-assembly, enabling matching sequences of bases to package genetic information in a duplex structure of paired helical strands, stabilized by the selective Watson–Crick (W-C) adenosine-thymine and guanine-cytosine base-pairing motif (2), and by the columnar stacking of the aromatic hydrocarbon nanosheets formed by the paired bases (3, 4). One of the great mysteries of evolution is how such a spectacular scenario first appeared in the universe.

Columnar molecular stacking is the key structural motif of the hierarchical self-assembly in chromonic lyotropic liquid crystals (LCs). This mode of molecular organization is driven by chromonic amphiphilicity, in which polycyclic aromatic hydrocarbon nanosheets, rendered soluble by hydrophilic substitutions (H bonding, polar, or ionic), associate hydrophobically into columns stabilized by stacking interactions (5–8). The similarity of such chromonic stacking to that of the bases in NAs has been noted (9, 10), where, in the NA case, the nanosheets consist of sets of mutually H-bonded bases. Such a two-step assembly motif leads to stacking interactions sufficiently strong to give chromonic columnar LC ordering in aqueous solution of, for example, guanosine nucleotide monomers which H-bond into planar tetramers (11), in mixtures of purine nucleobases and cyanuric acid, which H-bond into planar hexamers (12), and in heterocycle dimers (13).

Until now, stable duplex columnar ordering of NAs has required some degree of polymer linking of the stacking units, having been obtained as: (*i*) W-C duplexes (1) and columnar LCs (14) in NA polymers (Fig. 1 *D* and *I*); (*ii*) as W-C duplexes in mixtures of NA polymers and monomers (15–17); and (*iii*) as the chromonic columnar (COL) LC phase of ultrashort DNA and RNA oligomers (Fig. 1 *C* and *I*) (18–21). The latter are stabilized by the hierarchical self-assembly (HSA) of short complementary duplexes into linear aggregates via end-to-end attractive interactions (22–24), as a form of living polymerization (25–27), and

the collective condensation of these aggregates into distinct birefringent LC droplet domains (18, 20, 22). Within these phase-separated LC domains, the long-ranged LC order creates a self-determined equilibrium state that performs molecular selection and can greatly enhance the effective concentration and stability of end-to-end oligomer contacts (18, 23, 25–27). As a result, the combination of selectivity, molecular organization, and fluidity uniquely provided by the LC ordering strongly promotes, with the addition of the appropriate chemistry, the covalent ligation of the oligomers of neighboring stacked duplexes into longer complementary strands (28). Such polymerization, in turn, generically stabilizes LC order, enhancing it in a positive feedback system of HSA, selection, and synthesis, which has been termed "liquid crystal autocatalysis" (LCA), and proposed as a path toward the prebiotic appearance of sequence-directed self-assembling linear polymers (28, 29). However, starting LCA from even few-base-long oligomers or by relying on polymer templates is problematical in a prebiotic context, since such molecular families are already highly selected and entropically narrowed, and could not have previously appeared without effective positive feedback and molecular selection (30). This motivates our search for (COL) chromonic LC order in the form of phase-separated LC domains capable of molecular selection and LCA in mixed starting populations of more primitive molecular species, such as NA monobases. Such lyotropic LC ordering has not been found to date in NA monobase variants without H bonding or with H-bonding

Fig. 1. Lyotropic chromonic LC ordering of nucleoside triphosphates. (*A* and *B*) Chemical structures of dATP and dTTP and of a W-C hydrogen-bonded dATP/dTTP dimer showing its hydrophobic and hydrophilic domains. The hydrogen bonding reduces the solubility of the base pair, promoting the chromonic stacking of bases. (*C*) Uniaxial hexagonal COL LC phase of polymeric B-DNA. (*E*) COL LC phase formed by the self-assembly of anisotropic aggregates of duplex paired molecules. Such "chromonic" aggregation is driven by base pairing and stacking to shield flat hydrophobic molecular components from contact with water. (*G*) COL LC ordering of duplex NTP stacks, sketched for W-C duplexing, as indicated by the experiments. The sketch of the base pair indicates its chiral symmetry. The subcolumns of single bases have opposite $5' \rightarrow 3'$ directions, as in W-C NAs, making the two columns of phosphates and sugars structurally identical on average. The black dots indicate the helical stacking in the columns, forced by the bulky triphosphate groups, and the yellow arrows the 5' to 3' direction. (*D*, *F*, and *H*) Polarized transmission optical microscopic textures of the characteristic conformal domains of the COL liquid crystal phase of (*D*) long (900 bp) B-DNA (concentration, $c \sim 500$ mg/mL); (*F*) the self-cDNA hexamer 5'-CGACG-3' ($c \sim 800$ mg/mL); and (*H*) the dATP/dTTP mixture ($c \sim 900$ mg/mL). (*I*) Schematic illustration of a conformal domain of the COL phase showing the local geometry of columns of stacks of oligomeric NA, long NA, and NTP duplexed bases. The column packing permits bend deformation of the director field $n(r)$, the local column axis direction, but not splay, the key feature of COL ordering that produces its characteristic conformal domains. The COL birefringence $\Delta n = n_{\parallel} - n_{\perp}$ is negative because of the larger optical polarizability for polarization parallel to the planes of the bases. [Image width: 300 μm (*D*), 250 μm (*F*), 150 μm (*H*).]

motifs that favor only pairing, although their isotropic solutions exhibit features indicative of local (short-ranged) base pairing and stacking (31–39).

Results

Liquid Crystal Ordering of NTPs in Aqueous Solution. We report here the duplex [COL chromonic HSA and LC ordering of the mononucleoside triphosphates, chosen over other mononucleoside species because of their enhanced aqueous solubility and

potential for ligation provided by the triphosphate tail (40)]. The molecules investigated were the DNA mononucleoside triphosphates dATP, dCTP, dGTP, dTTP, and dUTP, and the RNA mononucleoside triphosphates rATP, rCTP, rGTP, rTTP, and rUTP. LC phase behavior in aqueous solution and the structure of the LC COL ordering was systematically studied by polarized transmission optical microscopy (PTOM) and X-ray diffraction (XRD) vs. temperature *T*, and NTP anion concentration, *c*(mg/mL) (*SI Appendix*).

Typical results for dATP/dTTP (AT) mixtures at a 1:1 molar ratio are shown for PTOM in Fig. 1*H* and for XRD in Fig. 2 and *SI Appendix*, Fig. S7 *B, D, and E*. The birefringent fan textures of the AT mixtures in Fig. 1*H*, which appear for $c > \sim 800$ mg/mL (NTP volume fraction $\varphi > \sim 0.6$) are characteristic of the lyotropic uniaxial COL ordering of long B-DNA (Fig. 1*D*) and of oligomeric B-DNA (Fig. 1*F*). The COL self-assembly in each case (Fig. 1 *E, C, and G*) allows only bend deformation of the director field *n*(*r*), taken to be along the stack axis (Fig. 1*I*). Absence of stripes parallel to *n*(*r*) indicated that the COL phase is uniaxial and therefore with the columns packed on a hexagonal lattice.

XRD is an effective probe of the structure of the partially ordered LC states of NA polymers (3, 41) and oligomers (18) in the regimes of high concentration, exhibiting peaks at wavevector $q_U = 2\pi/L_U$, giving the molecular base-stacking periodicity of $L_U \sim 3.3$ Å, and wavevector $q_h = 4\pi/\sqrt{3} \, a_h$, giving the intercolumn separation in the hexagonal lattice of the COL ordering (23 Å $< a_h <$ 33 Å). Fig. 2 *A and C* shows an example of this behavior for the COL LC phase of the self-complementary nanoDNA dodecamer 5′GCGCTTAAGCGC3′ (dD1). Fig. 2 *B and D* show that in the dATP/dTTP monomer mixtures, LC ordering is heralded in XRD by (*i*) a peak at $q_U \cong 1.88$ Å^{-1}, indicating that $L_U \cong 3.34$ Å, independent of concentration *c*, the same, within uncertainty, as that of B DNA; and (*ii*) a single peak reflection at $q_h \sim 0.3$ Å^{-1}, indicating that $L_U \cong 24$ Å, which depends on *c*. This measurement of L_U and a_h enables calculation of effective molecular volume (42) and geometry, which, in turn, enables the determination of the multiplicity of bases in the unit cell in the COL chromonic stacks, as detailed in *SI Appendix*, sections S2 and S3 and Fig. S8. The X-ray results indicate that the dATP/dTTP (AT) columns are unambiguously chromonic duplex stacks, which suggests the structural analogy presented in Fig. 1 *E, C, G, and I*, and associates the uniaxial hexagonal COL phase textures with duplex COL ordering in long DNA, DNA oligomers, and dATP/dTTP.

Textural observations of phase behavior were made on a variety of aqueous mixtures of dNTPs, as summarized in Figs. 3 and 4. Birefringent COL LC domains were observed in the 1:1 dCTP/dGTP (CG), and in 1:1:1:1 dATP/dCTP/dAGP/dTTP (ACGT) mixtures, exhibiting PTOM characteristics similar to those discussed for the AT mixtures in Fig. 1*H*. XRD analysis of the COL phase in 1:1 dCTP/dGTP mixtures confirm a structure of duplex columns formed by stacked pairs of bases, as for the dATP/dTTP mixtures. On the other hand, the solutions of binary dNTP mixtures AC, AG, CT, GT, and the solutions of the single dNTP bases A, C, G, and T, exhibited no COL ordering under the range of conditions studied [100 °C $> T > -20$ °C and $c < c_{max} = 1,350$ mg/mL, the maximum achievable NTP anion concentration (*SI Appendix*)]. Among these solutions the binary mixtures AC, CT, and the single-base solutions A, C, and T exhibited only the isotropic (ISO) phase. However, all mixtures containing dGTP (those in the dGTP row and dGTP column as indicated with blue font in Fig. 3*A*), including the ones with only G, showed, for sufficiently large concentration c_{dGTP}, domains having birefringence similar to that of COL domains but with uniform LC director orientation (Fig. 3*C*), indicating a phase with positional ordering of neighboring columns sufficient to suppress the COL phase column curvature. This phase is denoted as COL2. The dCTP/dGTP (1:1) mixtures exhibited both the COL and COL2 phases, with only the duplex COL domains found for $c < 800$ mg/mL, and the COL2 domains dominating at higher concentrations ($c > 850$ mg/mL). In addition to the COL peak, XRD of the 1:1 dCTP/dGTP mixtures show also a COL2 peak at smaller *q*, consistent with scattering from hexagonal COL lattice having unit-cell dimension of a stacking of G-quadruplexes, as observed previously in guanosine nucleotide solutions (*SI Appendix*, Figs. S7*F* and S8) (43). The 1:1:1:1 dATP/dCTP/dAGP/dTTP mixtures also exhibited COL phase textures at temperatures comparable to those of the AT and CG (*SI Appendix*, Fig. S15).

Several rNTP and rNTP/dNTP binary mixtures were also investigated at temperatures $T > -20$ °C, with the results summarized in the matrix of observations of the COL phase shown in *SI Appendix*, Fig. S16. Of note is the appearance of the COL phase in rCTP/rGTP and rATP/rTTP mixtures, but not rATP/rUTP mixtures. The latter absence is an observed break from the
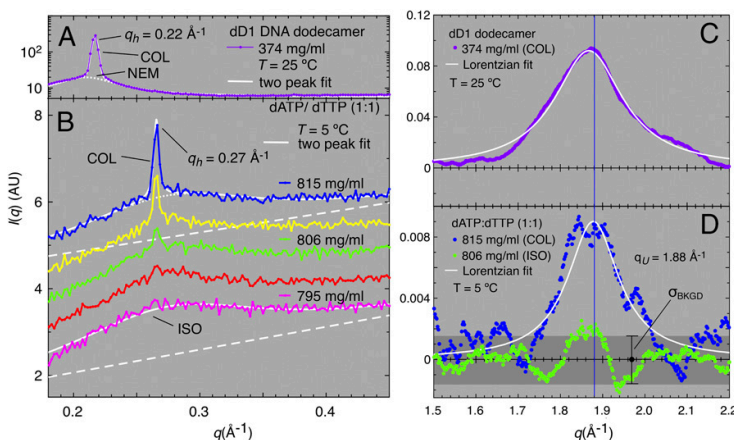
Fig. 2. X-ray structure factors *I(q)* showing the periodicities in the COL LC phase due to the packing of duplex columns of chromonic aggregates and base stacking within the columns. (*A*) The dD1 dodecamer, included for comparison purposes, forms chromonic COL stacks of blunt-ended duplexes, which, at $c =$ 374 mg/mL order into coexisting uniaxial hexagonal COL and uniaxial nematic (NEM) phases. (*B*) Coexisting COL and ISO phases versus concentration *c* in the dATP/dTTP (1:1) mixture. These data are taken along the black box in *SI Appendix*, Fig. S9*C*. The diffuse peak in the ISO phase indicates short-ranged aggregation and packing of columns. The ISO peak is similar to that found in chromonic phases of organic dyes (65). (*C and D*) X-ray structure factors *I(q)* at large *q*, showing the base-stacking periodicity in the chromonic aggregates forming COL LC phases. The base-stacking peak is at $q_U = 1.88$ Å^{-1} in both dD1 and the NTP mixtures, corresponding to $L_U = 3.34$ Å. (*B and D*) Both the column ordering and base-stacking peaks disappear upon entering the ISO phase (green). (*D*) σ_{BKGD} indicates the shot noise of the background level.

Fig. 3. ACGT matrix summarizing the search for LC phases in aqueous solutions of dNTPs and their mixtures, showing a significant dependence on choice of base. (*A*) The matrix colors indicate the most ordered phase to appear for each base-pair combination for $T > -20 °C$. The COL phase appears in NTP solutions at high concentration only for pairs exhibiting W-C complementarity. Macroscopic phase ordering was evaluated by PTOM and internal structure by XRD. The general pattern of COL ordering is the self-assembly of COL chromonic stacks of H-bonded bases and the packing of the columns into two-dimensionally hexagonally ordered arrays, where the distinction between duplex and quadruplex stacking is made using XRD (*SI Appendix*, Figs. S7 and S8). (*A* and *B*) Green shading indicates the binary solutions dATP/dTTP (AT) and dCTP/dGTP (CG), which exhibited uniaxial hexagonal duplex COL LC phases of chromonic stacks of base pairs. (*A* and *B*) Pink shading indicates the solutions which do not exhibit COL ordering, including the binary mixtures AC, AG, CT, and GT, and the single dNTP solutions A, C, G, and T. (*C*) Blue shading and the blue racetrack symbol indicate solutions containing dGTP, including G only, that exhibit, at higher concentrations of dGTP, domains of uniform internal orientation indicating a higher-ordered COL2 phase of chromonic stacks of H-bonded base quadruplexes. (*D*) Green shading indicates that the four-component equimolar mixtures, dATP/dCTP/dGTP/dTTP (ACGT), also exhibited COL domains at the concentrations studied (Fig. 4), probably due to CG phase separation and ordering. The appearance of duplex COL phases only for the binary mixtures AT and CG, which are the only mixtures of polymeric NA columns in solution to exhibit complementary W-C duplexing, is evidence for a W-C base-pairing motif in the COL LC phase. (Image width: 150 μm.)

dNTP W-C pattern of finding the COL in pairs that are complementary in oligomeric DNA and not finding it in those that are not complementary. However, this behavior was further explored in the matrix with the indicated group of ATP/UTP and ATP/TTP mixtures of dNTP and rNTPs. These results indicate that the LC COL ordering is more stable with TTP than with UTP, and is more stable with dNTPs than with rNTPs. Thus, dTTP makes COL phases with both dATP and rATP, whereas dUTP makes COL phases with dATP but not rATP, and rUTP makes COL phase with neither. This hierarchy is consistent with the temperatures T_u observed for thermal unbinding of DNA/RNA W-C homopolymer pairs: $(T_u)_{dAdT} > (T_u)_{rAdT} > (T_u)_{rArU}$ (44, 45).

Additional evidence for selective base pairing comes from the phase behavior of binary dNTP mixtures in mixtures of unbalanced relative mole fraction x and constant c. *SI Appendix*, Fig. S14 shows the x dependence of the ISO-COL melting temperature T_{IC} for dATP/dTTP at $c = 840$ mg/mL The depression of T_{IC} that is symmetric in $\delta x = x - 1/2$, the deviation from equimolarity, can be understood as a result of the contribution of the entropy of mixing to the ISO-COL transition free-energy difference.

Using this assumption and the melting entropy (46–48) gives a reasonable estimate of this depression (*SI Appendix*), indicating the selectivity of the phase separation between a stoi-

chiometrically balanced 1:1 COL LC domains and an A-rich isotropic phase.

W-C Selectivity. The core observation presented here is that long-range duplex COL LC ordering, characterized by base stacking and columns of dNTP base pairs, is found only in the AT and CG combinations, among all solutions of single A, C, G, and T nucleotide triphosphates and their binary mixtures. We refer to this pattern of LC ordering as "Watson–Crick selectivity" (W-C selectivity) since it matches the occurrence of W-C base pairing and duplex COL ordering of complementary NA polymer chains in dilute solutions. W-C selectivity so defined is also a feature of mixtures of nucleoside monomers with NA homopolymers, for example in the formation of duplexes of polyU with AMP, but not with CMP or IMP (15, 17), and in the formation of duplexes of polyC with GMP (16).

W-C selectivity in the NTPs indicates that these molecules have specific modes of base pairing and/or packing despite the enhanced freedom afforded by the absence of one or both polymer backbones. The diverse modes of NA pairing and stacking in the presence of backbones (reverse W-C, Hoogsteen, sugar edge, parallel chain, antiparallel chain, A-type, Z-type, duplex, triplex, quadruplex, etc.) are well documented. If chromonic COL LC polymorphism is any indicator (6–8), there is potentially even more variety possible in the absence of backbones, for example

Fig. 4. Compilation of relevant oligomer and monomer DNA LC phase diagrams, determined by PTOM and XRD (*SI Appendix*, Figs. S7 and S9 and *SI Appendix*), showing the high-temperature extent of LC ordering in: blunt-end self-complementary dodecamer duplexes (dD1); triphosphate-terminated self-complementary dodecamer duplexes (TPdD1); the shortest duplex forming NA oligomers previously reported (dATTA, dGCCG) (28); and several dNTP solutions. The blue line is the ISO-COL transition temperature in the dATP/dCTP/dGTP/dTTP (1:1:1:1) mixture (Fig. 2A and *SI Appendix*, Fig. S13 and *SI Appendix*). These LC phases are all from columns of duplex aggregates, except for the dCTP/dGTP (1:1) mixtures: at $c \sim 640$ mg/mL where LC ordering first appears; the COL phase is an assembly of duplex C-G columns, whereas for $c > 850$ mg/mL coexisting G-quadruplex columns appear, exhibiting stronger COL2 LC phase thermal stability (Fig. 2C and *SI Appendix*, Fig. S9).

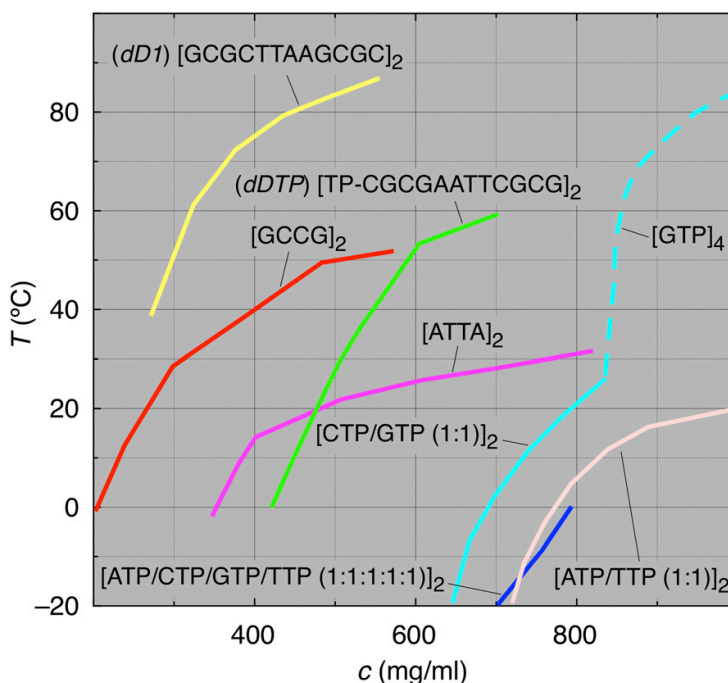flipping over every other molecule in a monomer stack (49), or every other pair in duplex stacks. W-C pairing is free-energetically favored in polymeric and oligomeric DNA, with the non–W-C arrangements coming into play as transient higher-energy states accessible via thermal fluctuation (50). Recalling the monomer/polymer systems (15–17), the extreme cooperativity of their hybridization transition (15) and FTIR studies show that the base-pairing selectivity of monomer/polymer mixtures that exhibit duplex stranding is W-C (16), indicating the persistence of the W-C pairing mode in the absence of one of the backbones (45). Therefore, although the freedom of organization of the stacks of monomers in the monomer/monomer systems should be greater than that of the polymer or monomer/polymer cases, W-C base pairing appears the best candidate to explain W-C selectivity in the totally backbone-free pairing and stacking condition of the NTPs, although our current structural data cannot distinguish among possible local structural motifs consistent with the measured L_U and a_h.

A schematic of a proposed W-C arrangement of the ATP/TTP COL phase is shown in Fig. 1G. The arrangement of the hydrogen-bonded bases and the antiparallel stacking of the 5′-3′ direction of the ribose rings is similar to that of polymer B DNA. In B DNA the base stacking is helixed because of the bulkiness of the ribose and the phosphodiester bond, and the convenient length of the latter. While the phosphodiester bond is absent in the NTP duplex stacks, the triphosphate group, having a molecular volume (~400 Å^3 per base) that cannot be accommodated in any simple vertical stacking, drives, along with the ribose, the tendency for adjacent base pairs to a nonzero relative

azimuthal orientation about the helix axis, z (Fig. 1G). The ribose chirality biases this relative orientation of bases: (*i*) toward helixing; and (*ii*) toward the antiparallel orientation of the two 3′-5′ ribose directions at a given level, making them symmetrically related by a π-rotation about x and therefore of the same handedness (Fig. 1G), which, as in polymer DNA, is required if the molecular conformations in the helical windings of the two phosphodiester/ribose chains are to be of the same average configurational structure. We thus propose a local structural organization very similar to that of the W-C of dilute polymeric duplex DNA, the double helix with triphosphates replacing the phosphodiester bonds of the polymer, as drawn schematically in Fig. 1G. In any case, whatever the details of the base-pairing geometry, the W-C selectivity indicates that that the COL LC ordering is a sensitive selection tool, capable on its own of stabilizing certain base-paired duplex DNA structures and destabilizing others in both the concentrated COL LC phase of monomers and the dilute duplex polymer chains.

Comparative Phase Behavior of NA Oligomers and Monomer Nucleotides. The optical and X-ray data enable the construction of LC phase diagrams for the two stoichiometrically balanced dATP/dTTP (1:1) and dCTP/dGTP (1:1) mixtures, shown in Fig. 4 and *SI Appendix*, Fig. S9, and their comparison with those of the four-component mixture, dATP/dCTP/dGTP/dTTP (1:1:1:1) and of a variety of previously studied complementary nanoDNA oligomers: DNA dodecamer 5′GCGCTTAAGCGC3′ (dD1); 5′ triphosphorylated DNA dodecamer TP-5′CGCGAATTCGCG3′ (TPdD1); and the DNA 4mers dATTA and dGCCG

(23). Fig. 4 is a summary of these phase diagrams, with the curves indicating the upper temperature limits of LC ordering. dD1 and TPdD1 are compared, which shows that oligomeric NAs can LC order in the presence of the extra charges and steric constraint of a terminal triphosphate, although with a reduced attractive interaction energy (*SI Appendix*). General trends evident in these data are (*i*) Solutions of single-bases NTPs do not exhibit LC ordering for $T > -20$ °C. (*ii*) Larger H bonding and stacking free energy makes the COL phase more stable (CG vs. AT). This effect is well-known in long NAs (22), has been confirmed in the short oligomers (23), and is evident in the dATTA and dGCCG curves. Remarkably, the same trend persists even in the chromonic HSA of COL stacks of dNTPs, a further, strong indication that pairing of the monobase NTPs in the COL LC phase takes place through a W-C geometry. (*iii*) The phase boundaries move to lower c and higher T with increasing oligomer length, indicating that LC phase stability is correspondingly enhanced by polymerization. Blunt-end self-cDNA oligomers in the length range $6 < n < 20$ duplexes exhibit similar behavior (18). (*iv*) For a given molecule LC stability is enhanced by increasing c, the behavior typical of chromonic LCs (6–8). (*v*) Triphosphate on the end of duplex chains destabilizes the COL phase. (*vi*) H bonding and stacking free energy is substantially larger in the G-quartet columns than for the duplex columns. (*vii*) LC stability extends to lower c in the GCCG and ATTA tetramers, systems with base overhangs. (*viii*) For the oligomers stabilized by blunt-end stacking and the NTPs, there is a relatively sharp cutoff concentration that increases strongly with decreasing oligomer length, below which LC ordering at low T is not possible.

The dCTP/dGTP 1:1 mixtures form duplex COL stacks at lower concentrations, with the quadruplex COL2 stacks starting to appear at higher concentrations, for $c > 820$ mg/mL This CG behavior is a result of the stacking units of the quadruplex columns being G tetramers, in which the bases each share four mutual H bonds instead of the two or three in W-C pairing, and the stacking internal energy is ~50% larger per multibase unit than in a CG W-C pair (43). Despite this, mixtures of melted polydC and polydG single strands anneal at low c to W-C duplexes upon cooling (51, 52), in part a result of the relatively higher entropic penalty at lower c for assembling groups of multiple Gs.

Discussion

The NTPs are shown here to be a family of molecules capable of self-assembling into stacked aggregates that long-range order into chromonic COL LC phases. The resulting assemblies feature strongly cooperative intermolecular interactions producing: (*i*) a first-order transition to the COL phase; (*ii*) the physical separation of distinct molecular species by LC droplets; (*iii*) 1:1 stoichiometry in the binary COL mixtures; and (*iv*) duplex stacking, all under W-C selectivity. These features of the NTP interactions bear a strong resemblance to those that characterize monomer/polymer (15, 16) and polymer/polymer NA mixtures (1). Thus, we find that the combination of selective H bonding, molecular packing, and nanosegregation driven by chromonic amphiphilicity (53) can generate the essential structural characteristics of polymeric duplex DNA by the hierarchical self-assembly of small molecules. Notably, the NTPs are also members of a molecular class that is potentially populated in the prebiotic era (54–57).

Fig. 4 shows that, For NTP pairs that are able to form the COL phase, polymerization strongly stabilizes LC ordering, with melting temperatures increasing from near freezing to near boiling as molecular length increases from monomers to dodecamers. The combination of these observations with our previous finding for DNA dodecamers that the LC ordering can itself promote NA polymerization (28) shows that chromonics which developed polymerization would have an evolutionary advantage in a quest to develop and grow LC order. The NTP system, with its potential for polymerization, thus exhibits conditions sufficient for LC autocatalysis (LCA), in which LC ordering stabilizes and expands itself by selecting molecules and catalytically promoting their ligation, in analogy to that in refs. 28 and 29.

The NTPs exemplify a class of trifunctional molecules (selective H bonding, chromonically amphiphilic, polymerizable) achievable via efficient synthetic pathways of relevant prebiotic chemistry (54–57). It is likely, however, that in a rich prebiotic synthetic environment, the available molecular species would be broadly distributed in structure and interaction characteristics. Narrowing such distributions requires molecular selection and positive feedback as the essential features that drive evolution in the presence of chemical processing. In LCA, the overall LC self-assembly itself is the "guiding hand" (30), with the LC ordered domains providing the selection and templating. The ACGT texture of Fig. 3*D* emphasizes the role of collectivity in this scenario. Cooperativity drives selection by phase separation, dividing space into sharply defined (phase-separated) volumes of ISO and COL LC order that exhibit differential molecular solubility, and therefore selection. If the COL phase can promote chemistry that enhances its stability, either from within as in ref. 28, or by enhancing other modes of catalysis, such as templating provided by inorganic (58) or lipid interfaces (59) for example, then the COL phase itself generates the positive feedback that drives enhanced H bonding and stacking, the essential requirement of LCA.

The molecular structures of the NA monomers and their multiplex COL ordering suggests in the context of LCA that the prebiotic emergence was via populations of one- and two-ring molecules (60, 61, 54–57), the most readily synthesized of the aromatic hydrocarbons and the most readily solubilized and functionalized by H bonding or polar groups. Such molecules do not form chromonic LCs in water as stacks of individuals, but, as the NTPs show, can be sufficiently amphiphilic to do so as multiplex columns. W-C selectivity shows unambiguously that such stacking is sensitive to subtle differences in the pairing energetics within a family of molecules that are relatively homogeneous in structure and suited to stack. However, for LCA to be a viable evolutionary process, or for that matter even to get started, chromonic COL LCs of multiplex stacking must be achievable in diverse multicomponent mixtures having molecular structural variations well beyond those of the realm of the nucleotides, a step that remains to be demonstrated.

LCA, as the basis of an "LC world" in which a combination of chromonic COL LC ordering and chemical activity selects and evolves molecular populations, has the potential to support the onset of an RNA world (62–64), by providing a feedstock of chains with diverse side groups capable of stacking and sequence-directed self-assembly, from both monomeric components and recycled deteriorated products.

Materials and Methods

LC phases of the dNTPs (dATP, dCTP, dGTP, dTTP) were produced from PCR-grade DNA nucleoside triphosphate solutions. DNA Dodecamers dD1 (5′ GCGCTTAAGCGC3′), dD2 (5′CGCGAATTCGCG3′), and 5′-Triphosphate (TPdD1, TP-5′CGCGAATT CGCG3′), were synthesized using an Äkta Oligopilot 100. PTOM was carried out on Nikon research microscopes equipped with polarizer and analyzer for the visualization of optic axis orientational textures. XRD experiments were performed on the SAXS/WAXS beamline 7.3.3 at the Advanced Light Source of Lawrence Berkeley National Laboratory in Berkeley, California. Observations were made using several different sample-cell preparations: (*i*) Evaporative cells, in which aqueous mixtures of NTPs were spotted on a glass substrate and covered with a glass slip or filled by capillarity between spaced plates, then allowed to evaporate while being monitored for the appearance of birefringent textures; (*ii*) Free-surface contact cells, in which water and vacuum-dried NTPs were brought into physical contact at a free surface under an oil seal; (*iii*) Flame-sealed capillaries, equilibrated and homogenized in the ISO phase by thermal cycling and centrifugal mixing; and (*iv*) Planar sealed cells of two glass slides spaced by a known, ~5–20-μm gap and filled with the sample of specified

concentration and surrounded by oil. Details of sample preparation and observation are available in *SI Appendix*.

1. Berg JM, Tymoczko JL, Gatto GJ, Jr, Stryer L (2015) *Biochemistry* (W.H. Freeman, New York), 8th Ed.
2. Watson JD, Crick FHC (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171:737–738.
3. Franklin RE, Gosling RG (1953) Molecular configuration in sodium thymonucleate. *Nature* 171:740–741.
4. Wilkins MHF, Stokes AR, Wilson HR (1953) Molecular structure of deoxypentose nucleic acids. *Nature* 171:738–740.
5. Jelley EE (1937) Molecular, nematic and crystal states of 1:1 '-diethyl-'V- cyanine chloride. *Nature* 139:631–632.
6. Attwood K, Lydon JE, Jones F (1986) The chromonic phases of dyes. *Liq Cryst* 1:499–507.
7. Lydon JE (1998) Chromonic liquid crystal phases. *Curr Opin Colloid Interface Sci* 3:458–466.
8. Lydon JE (2010) Chromonic review. *J Mater Chem* 20:10071–10099.
9. Mundy K, Sleep JC, Lydon JE (1995) The intercalation of ethidium bromide in the chromonic lyotropic phases of drugs and nucleic acids. *Liq Cryst* 19:107–112.
10. Lydon JE (2003) The DNA double helix—The untold story. *Liquid Crystals Today* 12:1–9.
11. Bonazzi S, et al. (1991) Four-stranded aggregates of oligodeoxyguanylates forming lyotropic liquid crystals: A study by circular dichroism, optical microscopy, and x-ray diffraction. *J Am Chem Soc* 113:5809–5816.
12. Li C, Cafferty BJ, Karunakaran SC, Schuster GB, Hud NV (2016) Formation of supramolecular assemblies and liquid crystals by purine nucleobases and cyanuric acid in water: Implications for the possible origins of RNA. *Phys Chem Chem Phys* 18:20091–20096.
13. Brunsveld L, Vekemans JAJM, Hirschberg JHKK, Sijbesma RP, Meijer EW (2002) Hierarchical formation of helical supramolecular polymers via stacking of hydrogen-bonded pairs in water. *Proc Natl Acad Sci USA* 99:4977–4982.
14. Livolant F, Levelut AM, Doucet J, Benoit JP (1989) The highly concentrated liquid-crystalline phase of DNA is columnar hexagonal. *Nature* 339:724–726.
15. Ts'o POP (1969) The hydrophobic-stacking properties of the bases in nucleic acids. *Ann N Y Acad Sci* 153:785–804.
16. Howard FB, Frazier J, Lipsett MN, Miles HT (1964) Infrared demonstration of two-and three-strand helix formation between poly C and guanosine mononucleotides and oligonucleotides. *Biochem Biophys Res Commun* 17:93–102.
17. Howard FB, Frazier J, Singer MF, Miles HT (1966) Helix formation between polyribonucleotides and purines, purine nucleosides and nucleotides. II. *J Mol Biol* 16:415–439.
18. Nakata M, et al. (2007) End-to-end stacking and liquid crystal condensation of 6 to 20 base pair DNA duplexes. *Science* 318:1276–1279.
19. Zanchetta G, Bellini T, Nakata M, Clark NA (2008) Physical polymerization and liquid crystallization of RNA oligomers. *J Am Chem Soc* 130:12864–12865.
20. Zanchetta G, Nakata M, Buscaglia M, Clark NA, Bellini T (2008) Liquid crystal ordering of DNA and RNA oligomers with partially overlapping sequences. *J Phys Condens Matter* 20:494214.
21. Bellini T, et al. (2012) Liquid crystal self-assembly of random-sequence DNA oligomers. *Proc Natl Acad Sci USA* 109:1110–1115.
22. Zanchetta G, Nakata M, Buscaglia M, Bellini T, Clark NA (2008) Phase separation and liquid crystallization of complementary sequences in mixtures of nanoDNA oligomers. *Proc Natl Acad Sci USA* 105:1111–1117.
23. Fraccia TP, et al. (2016) Liquid crystal ordering and isotropic gelation in solutions of four-base-long DNA oligomers. *ACS Nano* 10:8508–8516.
24. Saurabh S, et al. (2017) Understanding the origin of liquid crystal ordering of ultrashort double-stranded DNA. *Phys Rev E* 95:032702.
25. Kuriabova T, Betterton MD, Glaser MA (2010) Linear aggregation and liquid-crystalline order: Comparison of Monte Carlo simulation and analytic theory. *J Mater Chem* 20:10366–10383.
26. De Michele C, Bellini T, Sciortino F (2012) Self-assembly of bifunctional patchy particles with anisotropic shape into polymers chains: Theory, simulations, and experiments. *Macromolecules* 45:1090–1106.
27. Nguyen KT, Sciortino F, De Michele C (2014) Self-assembly-driven nematization. *Langmuir* 30:4814–4819.
28. Fraccia TP, et al. (2015) Abiotic ligation of DNA oligomers templated by their liquid crystal ordering. *Nat Commun* 6:6424, and erratum (2015) 6:463.
29. Fraccia TP, Zanchetta G, Rimoldi V, Clark NA, Bellini T (2015) Evidence for liquid crystal-assisted abiotic ligation of nucleic acids. *Orig Life Evol Biosph* 45:51–68.
30. de Duve C (2005) *Singularities: Landmarks on the Pathways of Life* (Cambridge Univ Press, New York).
31. Tso POP, Melvin IS, Olson AC (1963) Interaction and association of bases and nucleosides in aqueous solution. *J Am Chem Soc* 85:1289–1296.
32. Tso POP (1974) *Basic Principles in Nucleic Acid Chemistry* (Academic, New York), Vol 1.
33. Eimer W, Dorfmiller T (1992) Interaction of the complementary mononucleotides in aqueous solution. *J Phys Chem* 96:6801–6804.
34. Wissenburg P, Odijk T, Cirkel P, Mandel M (1995) Multimolecular aggregation of mononucleosomal DNA in concentrated isotropic solutions. *Macromolecules* 28:2315–2328.
35. Mariani P, et al. (2009) Small angle X-ray scattering analysis of deoxyguanosine 5'-monophosphate self-assembling in solution: Nucleation and growth of G-quadruplexes. *J Phys Chem B* 113:7934–7944.
36. Schweizer MP, Broom AD, Ts'o POP, Hollis DP (1968) Studies of inter- and intramolecular interaction in mononucleotides by proton magnetic resonance. *J Am Chem Soc* 90:1042–1055.
37. Raszka M, Kaplan NO (1972) Association by hydrogen bonding of mononucleotides in aqueous solution. *Proc Natl Acad Sci USA* 69:2025–2029.
38. Spindler L, Drevensek-Olenik I, Copic M, Cerar J, Skerjanc J (2004) Dynamic light scattering and 31P NMR study of the self-assembly of deoxyguanosine 5'-monophosphate: The effect of added salt. *Eur Phys J E Soft Matter* 13:27–33.
39. Kilchherr F, et al. (2016) Single-molecule dissection of stacking forces in DNA. *Science* 353:5508.
40. Rohatgi R, Bartel DP, Szostak JW (1996) Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J Am Chem Soc* 118:3332–3339.
41. Strey HH, Parsegian VA, Podgornik R (1999) Equation of state for polymer liquid crystals: Theory and experiment. *Phys Rev E* 59:999–1008.
42. Nadassy K, Tomás-Oliveira I, Alberts I, Janin J, Wodak SJ (2001) Standard atomic volumes in double-stranded DNA and packing in protein–DNA interfaces. *Nucleic Acids Res* 29:3362–3376.
43. Davis JT (2004) G-quartets 40 years later: From 5'-GMP to molecular biology and supramolecular chemistry. *Angew Chem Int Ed Engl* 43:668–698.
44. Riley M, Maling B, Chamberlin MJ (1966) Physical and chemical characterization of two- and three-stranded adenine-thymine and adenine-uracil homopolymer complexes. *J Mol Biol* 20:359–389.
45. Felsenfeld G, Miles HT (1967) The physical and chemical properties of nucleic acids. *Annu Rev Biochem* 36:407–448.
46. Klump HH, Völker J, Maeder DL, Niermann T, Sobolewski CHM (1991) Conformational changes in nucleic acids/chromatin structure. *Thermochim Acta* 193:391–415.
47. Gruenwedel DW (1975) Salt effects on the denaturation of DNA. IV. A calorimetric study of the helix-coil conversion of the alternating copolymer poly[d(A-T)]. *Biochim Biophys Acta* 395:246–257.
48. Vesnaver G, Breslauer KJ (1991) The contribution of DNA single-stranded order to the thermodynamics of duplex formation. *Proc Natl Acad Sci USA* 88:3569–3573.
49. Chami F, Wilson MR (2010) Molecular order in a chromonic liquid crystal: A molecular simulation study of the anionic azo dye sunset yellow. *J Am Chem Soc* 132:7794–7802.
50. Yang C, Kim E, Pak Y (2015) Free energy landscape and transition pathways from Watson-Crick to Hoogsteen base pairing in free duplex DNA. *Nucleic Acids Res* 43:7769–7778.
51. Gray DM, Bollum FJ (1974) A circular dichroism study of poly dG, poly dC and poly dG: dC. *Biopolymers* 13:2087–2102.
52. Hwang JS, et al. (2002) Electrical transport through 60 base pairs of poly(dG)-poly(dC) DNA molecules. *Appl Phys Lett* 81:1134–1136.
53. Attwood TK, Lydon JE, Hall C, Tiddy GJT (1990) The distinction between chromonic and amphiphilic lyotropic mesophases. *Liq Cryst* 7:657–668.
54. Benner SA, Kim HJ, Carrigan MA (2012) Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. *Acc Chem Res* 45:2025–2034.
55. Ruiz-Mirazo K, Briones C, de la Escosura A (2014) Prebiotic systems chemistry: New perspectives for the origins of life. *Chem Rev* 114:285–366.
56. Becker S, et al. (2016) A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* 352:833–836.
57. Powner MW, Sutherland JD, Szostak JW (2010) Chemoselective multicomponent one-pot assembly of purine precursors in water. *J Am Chem Soc* 132:16677–16688.
58. Ertem G, Ferris JP (1997) Template-directed synthesis using the heterogeneous templates produced by montmorillonite catalysis. A possible bridge between the prebiotic and RNA worlds. *J Am Chem Soc* 119:7197–7201.
59. Toppozini L, Dies H, Deamer DW, Rheinstädter MC (2013) Adenosine monophosphate forms ordered arrays in multilamellar lipid matrices: Insights into assembly of nucleic acid for primitive life. *PLoS One* 8:e62810.
60. Cafferty BJ, Hud NV (2015) Was a pyrimidine-pyrimidine base pair the ancestor of Watson-Crick base pairs? Insights from a systematic approach to the origin of RNA. *Isr J Chem* 55:891–905.
61. Patel BH, Percivalle C, Ritson DJ, Duffy CD, Sutherland JD (2015) Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat Chem* 7:301–307.
62. Gilbert W (1986) Origin of life: The RNA World. *Nature* 319:618.
63. Joyce GF (2002) The antiquity of RNA-based evolution. *Nature* 418:214–221.
64. Robertson MP, Joyce GF (2012) The origins of the RNA world. *Cold Spring Harb Perspect Biol* 4:003608.
65. Park HS, et al. (2008) Self-assembly of lyotropic chromonic liquid crystal sunset yellow and effects of ionic additives. *J Phys Chem B* 112:16307–16319.

Paper 3

Phase separations, liquid crystal ordering and molecular partitioning in mixtures of PEG and DNA oligomers.

 I have contributed to this work as follows:

– Paper writing;

– Characterization of PEG-RNA phase diagram.

Taylor & Francis
Taylor & Francis Group

INVITED ARTICLE

Check for updates

Phase separations, liquid crystal ordering and molecular partitioning in mixtures of PEG and DNA oligomers

Simone Di Leo[a], Marco Todisco[a], Tommaso Bellini[a] and Tommaso P. Fraccia[a,b]

[a]Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università di Milano, Segrate (MI), Italy; [b]Dipartimento di Scienze Umane e Promozione della Qualità della Vita, Università Telematica San Raffaele di Roma, Roma (RM), Italy
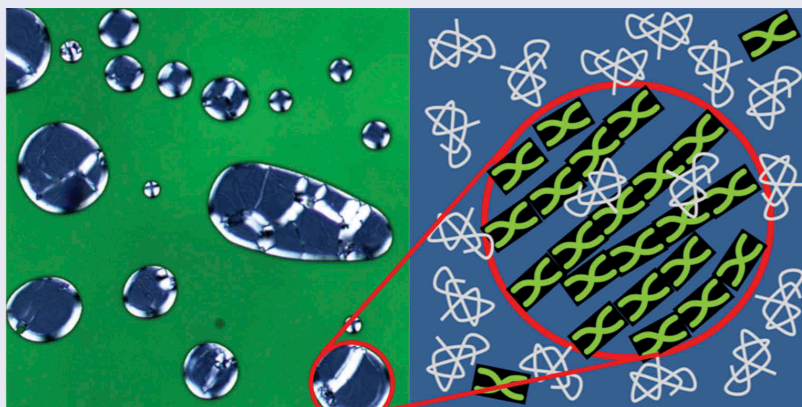
ABSTRACT

Liquid crystals (LCs) ordering of DNA and RNA oligomers relies on the presence of inter-duplex end-to-end attraction, driving the formation of linear aggregates. Such interactions are gauged, at a macroscopic level, by the osmotic pressure at the isotropic-nematic and nematic-columnar phase transitions. We studied aqueous solutions of PEG and DNA duplex-forming oligomers, finding that there is a wide range of concentrations in which these mixtures phase separate into coexisting PEG-rich and DNA-rich phases, the latter being either in the isotropic state or ordered as a nematic or columnar LC. We determined the phase diagram in mixtures of PEG and DNA duplexes with different terminal motifs – blunt ends, sticky overhangs, aggregation-preventing overhangs – and measured the partitioning of the species in the coexisting phases. On this basis, we determined the osmotic pressure as a function of the DNA concentration across the phase diagram. We compared the equation of state obtained in this way with both the Carnahan–Starling equation of state for hard spheres and with the pressure predicted by computer simulations of a system of aggregating cylinders. We obtain a good agreement between experiments and simulations, and end-to-end attraction energies of the order of 6 kcal/mol, a bit larger than expected, but still in agreement with the current models for DNA-DNA interactions.

1. Introduction

Concentrated aqueous solutions of DNA oligomers paired in duplexes have been reported to order into liquid crystal (LC) phases [1–4]. This tendency towards LC ordering is quite robust, being found in a variety of systems, ranging from well-paired ('blunt ended') duplexes [1], to duplexes with mutually interacting ('sticky') overhangs [2], to duplexes including various degrees of defects, to structures alternating double and single strands [5], to fully random sequences [6]. LCs have been found in longer (up to 20 bases) and shorter (down to 4 bases) DNA sequences [3,4]. Recently, LCs have been observed even in solutions of single nucleobases [7], a finding that demonstrates that stacking attractive forces [8] are the main driving force of LC ordering. Despite being at the core of DNA stability, stacking forces are still a matter of investigation [9], since both their fundamental nature and their strength

are not fully understood. In particular, stacking forces between independent duplexes have been quantified only via atomistic simulations [10].

The role of end-to-end attractions in the LC ordering of DNA duplexes has been studied via coarse-grained models. Specifically, LC ordering has been predicted in suspension of cylindrical monomer having steric repulsion and attractive interactions between their terminal bases only, which were investigated via computer simulations [11–14]. This is summarised in Figure 1 in which we plot the volume fraction (ϕ) – pressure (P) equation of state obtained from Monte Carlo simulations presented in Ref. [11]. for a system of hard cylinders ($L/D = 2$), with various choices of base-to-base interaction strength E, $0 < E < 12\ k_BT$. While at low pressure the system is an isotropic fluid (ISO), as concentration and pressure rise it may develop nematic LC (N), columnar LC (COL) and columnar crystal (C2) phases, depending on the interaction strength. As expected, the pressure at which the ISO-LC transition occurs decreases with increasing
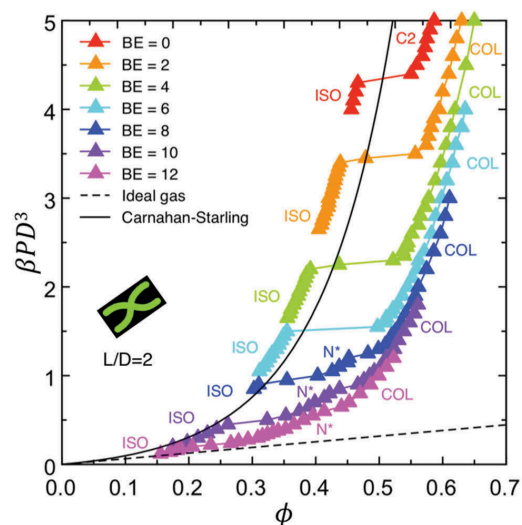
interaction strength. Thus, the measure of the pressure of the phase transition, which in the experimental systems would be the osmotic pressure (Π) of the DNA solution, enables estimating the strength of the inter-cylinder attractive interactions. In Figure 1 we also plot, as a simpler references: $\Pi_{CS}(\phi)$, the Carnahan–Starling (CS) equation of state for hard spheres [15], which lays close to the ISO branches of the simulation (black line), and the equation of state of the ideal gas (dashed line).

Experimental evaluation of DNA inter-duplex interactions were elaborated before by a quantitative analysis of the ISO-LC phase boundary [3,4,7] based on the predictions of the aggregation model [13].

In other systems of linear aggregating molecules, stacking energies have been evaluated by sensing the length distribution of the aggregates [16]. This has been done to study the molecular aggregation in chromonic LCs both with diffusion nuclear magnetic resonance (NMR) [17] and X-ray diffraction [18]. Such methods are however difficult to extent to the aggregation of DNA oligomers, since in this case each monomer has already an internal periodicity which would interfere with the detection of the duplex–duplex interaction.

In this work, we focus on measuring the pressure to enable a different, and possibly more direct route to evaluating the inter-duplex interaction.

A simple strategy to measure the osmotic pressure of DNA oligomers in solution is to exploit the phase separation that is observed when the flexible hydrophilic polymer polyethylene glycol (PEG) is solubilised in the same sample. PEG is widely used to produce aqueous two-phase systems (ATPS) in mixtures with salt, as potassium phosphate [19] and sodium succinate [20], or polymers, as dextrane [21,22], and to compartmentalise or precipitate biomolecules, as proteins [23–26] or nucleic acids [27]. The use of PEG as a reference osmolyte is a standard approach to study interactions and conformations of proteins [28–31] and viruses [32] in condensed solutions, since they readily phase separate into a protein (or virus) rich phase coexisting with a PEG-rich solution.

When the partitioning of the two solutes is total, or nearly so, the osmotic pressure of PEG, which has been fully characterised [33,34], equals the osmotic pressure of the coexisting phase. By exploiting this notion, PEG has been used as crowding agent to study the phase diagram and the osmotic pressure of the lyotropic chromonic LC sunset yellow [35]. Since PEG also phase separates with DNA, it has been used as a controlled osmolyte to measure the osmotic pressure of counterion condensed long DNA double helices [36], and across the isotropic-cholesteric and cholesteric-line

Figure 1. (Colour online) Equation of state for a system of sticky cylinders of diameter D and axial ratio $L/D = 2$, expressed in dimensionless pressure, βPD^3, where $\beta^{-1} = k_BT$, vs. volume fraction ϕ. The end-to-end binding energy E ranges from 0 to 12 k_BT. Data obtained from Monte Carlo simulations, adapted with permission from Ref. [11]. Coloured triangles correspond to different energies, see legend, and phases are as indicated by coloured labels, ISO = isotropic, N* = cholesteric, COL = columnar and C2 = columnar crystal. Lines connecting the symbols are drawn to help the identification of the first-order phase transitions, which take place at constant pressure. Solid black line is the Carnahan–Starling equation of state for hard spheres and dashed black line is the ideal gas equation.

hexatic phase boundary in solutions of long double-stranded DNA [37,38].

Phase separation is also observed in mixed solutions of PEG and DNA oligomers [39–41] as well as in PEG and RNA oligomers even for chains as short as 4 bases (unpublished data). This phenomenon has not been so far studied in detail. Phase separation of polymeric species is quite common, a phenomenon typically described in terms of the Flory–Huggins model, according to which the degree of polymerisation is a key factor in the reduction of the entropy of mixing. When the degree of polymerisation of one of the species is small, as in the case of PEG polymers and short DNA or RNA oligomers, the partitioning of the species might become partial.

When the phase separation between PEG and proteins or long DNA is 'strong', with a nearly complete partitioning of the two molecular species between the two phases, the determination of the osmotic pressures is straightforward both because (i) the PEG concentration in the PEG-rich phase directly gauges the overall osmotic pressure and because (ii) the PEG concentration in the PEG-rich phase is simply determined from the initial concentrations and the volume fractions of the coexisting phases. When instead there is a partial partitioning between the species, both these steps must be reconsidered. For clarity, let's call phase 1 the PEG-rich phase; phase 2 the DNA-rich phase; and $c_{DNA,1}$, $c_{DNA,2}$, $c_{PEG,1}$ and $c_{PEG,2}$ the concentrations of DNA and PEG in phase 1 and phase 2, respectively (see Figure 2). The presence of the minority components, $c_{DNA,1}$ and $c_{PEG,2}$ prevents from simply attributing all PEG to phase 1 and all DNA to phase 2, as generally done when the osmotic pressure is measured via phase separation with PEG [36]. This issue will be discussed in Section 3, where we describe the experimental determination of all concentrations. At the same time, the presence of the minority components affects the osmotic pressure, as discussed in Section 4.

2. Phase diagrams in mixtures of PEG and DNA oligomers

We extensively investigated aqueous mixtures of DNA oligomers and PEG having overall concentrations of
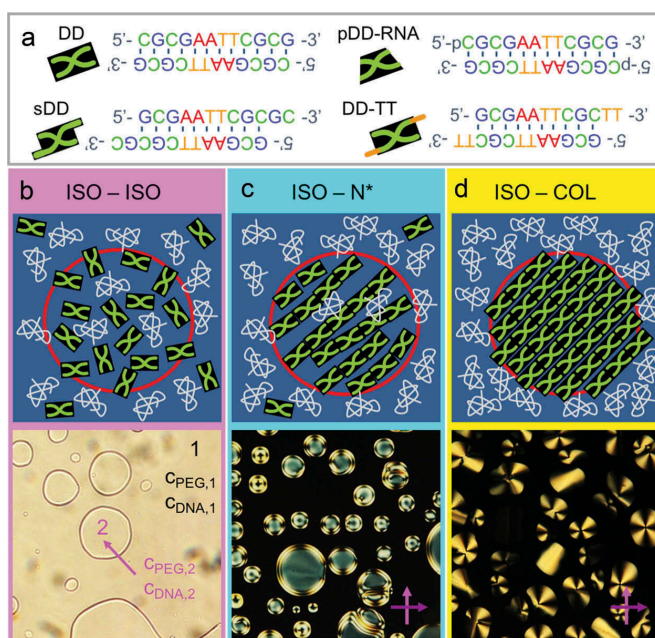
Figure 2. (Colour online) (a) Sketches and sequences of the DNA and RNA duplexes considered in the present work: blunt-end DD and RNA 5′phosphate DD, sticky-end sDD and non-sticky-end DD-TT, with corresponding sequences. (b–d upper panels) Pictorial representation of ATPS in mixtures of duplexes and PEG, sketched as grey bundles. Depending on concentrations, the system separates into a PEG-rich phase (phase 1) and a DNA-rich phase (phase 2), where $c_{PEG,1}$, $c_{DNA,1}$ and $c_{PEG,2}$, $c_{DNA,2}$ are the concentrations of PEG and DNA in phase 1 and phase 2, respectively. Upon increasing PEG and DNA concentrations, the partitioning of the two species becomes more effective, with progressively decreasing amount of DNA in phase 1 and PEG in phase 2. (b–d lower panels) Bright field micrograph of ISO-ISO phase separation (b), polarised optical micrographs of ISO-N* (c) and ISO-COL (d) phase separations.

c_{DNA} and c_{PEG}, respectively. We considered three different sequences of DNA and two PEG molecular weights, 8 and 20 kDa. DNA sequences have been chosen to provide duplexes with end-to-end interactions of different kind and strength, as sketched in Figure 2(a). In particular, the 'Dickerson Dodecamer', DD (5′-CGCGAATTCGCG-3′), is a self-complementary sequence that forms blunt-end duplexes that experience a mild end-to-end attraction because of the stacking interaction of their terminal paired bases. The DD duplex structure, of B-DNA type, has been well characterised by previous works [42]. We also considered modified version of DD, the 'shifted Dickerson Dodecamer', sDD, (5′-GCGAATTCGCGC-3′), obtained by relocating the first Cytosine nucleobase to the opposite (−3′) end. sDD is thus self-complementary only in a subset of 10 bases, and thus forms duplexes with 2-base-long GC overhangs at the 3′-end. Such unpaired bases (or 'dangling ends') can pair with their counterpart of other sDD duplexes, thus acting as sticky-ends which drive the inter-duplex interaction. Both these sequences are composed of 12 nucleotides, therefore the length, L, of the duplexes is $L \approx 4.1$ nm. Given that the DNA helix diameter is approximately $D \approx 2$ nm, the axial ratio of DD and sDD duplexes is $L/D \approx 2$. Both DD and sDD c_{DNA}-T phase diagram has been already investigated, showing the formation of cholesteric (N*), uniaxial columnar (COL) and crystalline columnar (C2) LC phases [1,2,7,43].

The third DNA sequence, DD-TT (5′-CGCGAATTCGCG-TT-3′), is a 14mer composed by the same DD sequence with the addition of two 3′-terminal thymine nucleobases. The almost negligible T-T pairing affinity lowers inter-duplex interaction, in turn preventing the formation of LC, while the axial ratio, $L/D \approx 2.4$, is kept of the order of the other duplexes. In addition, we also tested the RNA analogue of DD, pDD-RNA, which has the same sequence of DD except having uraciles instead of thymines, to test the universality of PEG–nucleic acid oligomers phase separation. Indeed, while DNA is most commonly a B-form helix, RNA duplexes have an A-form helical structure, in which the base pairs are tilted with respect to the helical axes. Because of this different helical geometry, the axial ratio of pDD-RNA duplexes is $L/D \approx 1.5$. Despite such structural differences and the presence of an additional phosphate at each 5′-terminus, similar LC behaviour has been observed either for pDD-RNA alone [44] or mixed with PEG (unpublished data).

We have here investigated four different systems: DD + PEG 8 kDa, 0.01 M HEPES buffer pH 7.5 and 300 mM NaCl; pDD-RNA + PEG 8 kDa, 0.01 M HEPES buffer pH 7.5 and 60 mM MgCl₂; sDD + PEG 20 kDa with no buffer nor added salt; DD-TT + PEG 20 kDa, with no buffer nor added salt.

The different choices were related to the availability and storage condition of the samples. For each of these systems and for every choice of c_{DNA} and c_{PEG}, samples were prepared in flat glass cells and observed in polarised optical microscopy (see Section 7).

In all systems, at low enough PEG concentration, the solutions are homogeneous. Upon increasing PEG and DNA (or RNA) concentrations, phase separations are eventually observed in all systems, that split into a PEG-rich isotropic phase and a DNA-rich phase that can be either isotropic or liquid crystalline. We have thus classified the phase separations on the basis of the ordering of the DNA-rich phase: ISO-ISO, when DNA phase is isotropic, Figure 2(b); ISO-N*, when DNA phase shows N* ordering, Figure 2(c); ISO-COL, when DNA phase shows COL ordering, Figure 2(d) and ISO-C2, when DNA phase shows C2 ordering. Phase separation follows a mechanism of nucleation and growth process [39]. After the growth, taking place in a few minutes, and an equilibration period of 1–2 days, DNA-rich domains have sizes ranging from few to hundreds of micrometers. Their shape is typically circular at low PEG content, i.e. lower average viscosity, or branched at high PEG content.

Figure 3 summarises the observed phase behaviour for the four systems, with the use of ternary phase diagrams. Each point on the ternary diagram uniquely indicates its DNA, PEG and water mass composition (w/w). This is illustrated in the construction in Figure 3 (d) for point 'a', whose composition is 0.4 w/w DNA, 0.1 w/w PEG and 0.5 w/w water. All experimental observations are reported in the ternary diagrams, with black squares for isotropic mixtures, magenta pentagons for ISO-ISO phase separations, blue dots for ISO-N* phase separations, red triangles for ISO-COL phase separations and green diamonds for ISO-C2 phase separations. The empty symbols on the zero PEG axis are the experimental points describing the phase diagrams of DNA solutions with no PEG. Dotted lines, parallel to the water axes, indicate lines of constant DNA and dashed lines, parallel to the DNA axes, are lines of constant PEG. By comparing the four diagrams, we can determine the general behaviour of oligoDNA-PEG mixtures: (i) at very low DNA concentration, a large PEG w/w is required to induce phase separation (see line 1 in Figure 3(c)); (ii) upon increasing the amount of DNA, the PEG w/w threshold for phase separation lowers (see lines 2 and 3 in Figure 3 (c); (iii) there is a range of DNA concentrations at which all the phases are obtained by increasing PEG w/w. This is the case of the samples laying on line 4 in Figure 3(a) (c_{DNA} = 200 g/l) along which we have found ISO, N*, COL and C2 phases; (iv) at constant
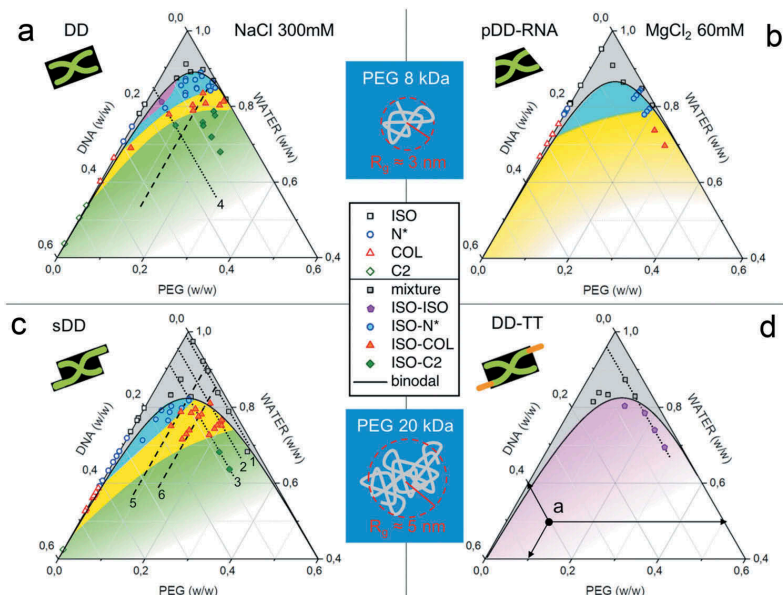
Figure 3. (Colour online) Ternary diagrams of the four investigated mixtures in unities of mass fraction w/w (i.e. w_{DNA}/w_{TOT}, w_{PEG}/w_{TOT} and w_{WATER}/w_{TOT}, where $w_{TOT} = w_{DNA} + w_{PEG} + w_{WATER}$. (a) DD/PEG-8 kDa/NaCl-300 mM, (b) pDD-RNA/PEG-8 kDa/MgCl₂-60 mM, (c) sDD/PEG-20 kDa/water and (D) DD-TT/PEG-20 kDa/water. Point 'a' in panel (d) and arrows provide an example for the reading of the system composition (0.4 DNA w/w – 0.1 PEG w/w – 0.5 water w/w). Symbols within the diagram indicate the initial samples preparations and their colours indicate the type of separation observed, according to the legend. Symbols laying on the axes of 0 PEG w/w indicate the phase diagrams of the pure DNA/RNA solution at room temperature. Solid black lines are the binodal lines, determined as described in the text, separating the region of homogeneous solution (grey shading) and the phase-coexistence region (violet, blue, yellow and green shading for ISO-ISO, ISO-N*, ISO-COL and ISO-C2 coexistence, respectively). Dotted lines indicate constant DNA preparations and dashed lines indicate constant PEG preparations.

PEG w/w, phase separation is found above a threshold DNA w/w (see line 5 in Figure 3(c)), which decreases as PEG w/w increases (see line 6 in Figure 3(c)); and (v) at high PEG w/w only highly ordered DNA phases survive, as C2 and amorphous crystals.

Having examined a large set of samples having different c_{DNA} and c_{PEG}, we could approximately determine the binodal line as interpolation between the concentrations of samples that do and do not phase separate. Such interpolated concentrations are plotted in Figure 4 as symbols. To provide an analytic expression to the binodal curves, we fitted the experimental data with the function $y = m_1 \cdot \exp(m_2 \cdot \sqrt{x} + m_3 \cdot x^3)$ empirically proposed to describe the binodal curve of PEG – potassium phosphate solutions [23], where y = PEG w/w and x = DNA w/w. Since our data are gathered in the region of low DNA w/w, i.e. small x, the parameter m_3 has little relevance. We thus simplified the functional expression by setting $m_3 = 0$, obtaining

$$c_{PEG} = m_1 \cdot \exp(m_2 \cdot \sqrt{c_{DNA}}) \tag{1}$$

The binodal lines obtained in this way are shown in Figure 4 as dashed lines and in the diagrams of Figure 3 as black lines. By comparing the resulting binodal lines, it can be observed that DD and pDD-RNA separate more easily than sDD and DD-TT since their binodal is shifted towards lower DNA and PEG content. This difference is probably due to the presence of added monovalent and divalent salts in the DD and pDD-RNA systems, respectively, in line with previous observations indicating that the formation of ATPS in PEG mixtures is promoted by higher ionic strength [23]. Moreover, the presence of cations has a strong effect on DNA: they generally reduce the lateral electrostatic repulsion between helices by screening the negative charges of the DNA backbone phosphates [45], while divalent and multivalent cations, such as Mg^{2+} and spermine, favour DNA condensation [46–49].
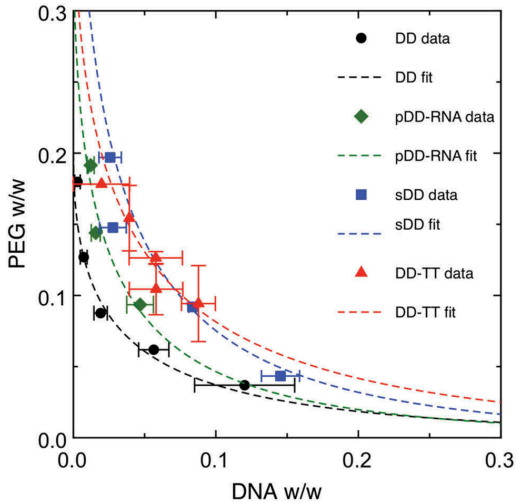
Figure 4. (Colour online) Experimental determination of the binodal lines in PEG w/w vs. DNA w/w diagram. Points are obtained as interpolation of the compositions of pairs of samples laying on the opposite sides of the binodal line. Error bars indicate the distance in c_{PEG} and c_{DNA} between the pairs. Dashed lines are the binodal obtained by fitting the points with the function $c_{PEG} = m_1 \cdot \exp\left(m_2 \cdot \sqrt{c_{DNA}} + m_3 \cdot c_{DNA}{}^3\right)$ from Ref. [23], simplified by assuming $m_3 = 0$. Binodals closest to the axes indicate stronger partitioning of PEG and DNA between the phases.

3. Partitioning of DNA and PEG

The very existence of measurable binodal curves indicates that for the majority of the preparations the partitioning of PEG and DNA in phase 1 and 2 is not complete. By defining ϕ_1 and ϕ_2 as the fractions of the total volume occupied by phase 1 and 2, respectively, mass conservation of PEG and DNA implies that

$$c_{PEG} = c_{PEG,1} \cdot \phi_1 + c_{PEG,2} \cdot \phi_2$$

$$c_{DNA} = c_{DNA,1} \cdot \phi_1 + c_{DNA,2} \cdot \phi_2 \qquad (2)$$

where $\phi_1 + \phi_2 = 1$. Since coexisting phases lay on the binodal curve, Eq. 1 holds for both phases, i.e. $c_{PEG,1} = m_1 \cdot \exp\left(m_2 \cdot \sqrt{c_{DNA,1}}\right)$ and $c_{PEG,2} = m_1 \cdot \exp\left(m_2 \cdot \sqrt{c_{DNA,2}}\right)$. This set of equations enables, at least in principle, determining the concentrations in both phases on the basis of c_{DNA} and c_{PEG} once ϕ_1 and ϕ_2 are known.

We experimentally measured ϕ_1 and ϕ_2 by adding a fraction of fluorescent PEG (FITC labelled PEG 5 kDa) to the PEG stock solutions, at the 1:100 stoichiometric ratio between fluorescent PEG and the unlabelled PEG. Microscope images of the total sample area were collected

and analysed to extract the volume fractions of the two phases, as shown in Figure 5(a–c), (see Section 7). In practice, the measured values of ϕ_1 and ϕ_2 do not enable to extract reliable values of the whole set of concentrations probably because the approximation on the 'tails' of the binodal curve at large DNA content is too rude.

More direct information is obtained by analysing the fluorescent emission of fluorinated PEG, as described in Figure 5. The two peaks' shape of the
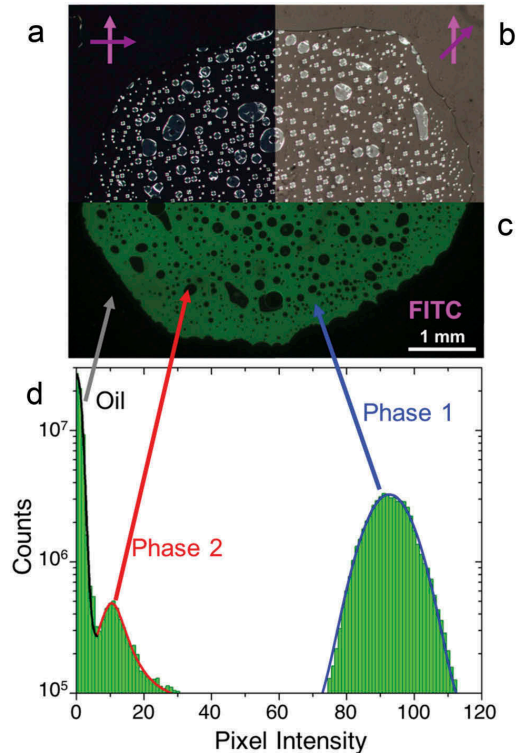
Figure 5. (Colour online) Low magnification (2×) optical microscopy images of ISO-N* phase separation in DD/PEG-8 kDa/NaCl-300 mM with PEG w/w = 0.082 and DNA w/w = 0.089, with crossed polarisers (a), partially uncrossed polarisers (b) and fluorescent emission (c). Fluorescent PEG 5 kDa, tagged with FITC fluorophore, has been mixed to untagged PEG at 1:100 stoichiometric ratio. Fluorescence signal enables direct measurement of the volume fractions. In panel C, $\phi_1 = 0.80 \pm 0.05$ and $\phi_2 = 0.20 \pm 0.05$. (D) Histogram of the fluorescence intensity measured on the 8-bit greyscale image (black = 0, white = 255). The three peaks are associated to the fluorescent emission of: the oil background area (black contour peak and grey arrow), the PEG-rich phase 1 (blue contour and blue arrow) and the DNA rich phase 2 (red contour and red arrow). The ratio of the mean intensity of phase 1 and phase 2, after subtraction of the oil background, gives the $c_{PEG,1}/c_{PEG,2}$ ratio, a value necessary to determine the composition of the two phases (see the text).

histogram of the fluorescence intensity (Figure 5(c)) shows that while the fluorescence signal mostly comes from phase 1, rich in fluorescent PEG, there is still non-negligible signal from phase 2. The ratio of the average intensity measured in the two phases, $r = \bar{I}_{PEG,2}/\bar{I}_{PEG,1}$, can be considered a direct measurement of $c_{PEG,2}/c_{PEG,1}$. Knowing $c_{PEG,1}$ we calculate $c_{DNA,1}$ from the expression of the binodal line, and then $c_{DNA,2}$ is obtained from DNA mass conservation $c_{DNA,2} = (c_{DNA,0} - c_{DNA,1} \cdot \phi_1)/\phi_2$.

By comparing the two approaches, we opted to use the direct measurement of $c_{PEG,2}/c_{PEG,1}$ for the samples with larger DNA content, while we used the approach based on the analytical expression of the binodal line at low DNA content, where the majority of the available experimental data lies, making the determination of the lines quite robust. Anyway, the two methods are highly compatible, confirming that analytical expression of the binodal line is a good description of the actual behaviour of the system. Based on this analysis, we could draw the tie lines, as shown in Figure 6 for a selection of the experimental points of the DD and sDD systems.

The main source of error is the measurement of the volume fractions ϕ, which we estimated having the sensibility $\approx 0.5\%$ of the total volume of the cell. Thus, the measurements of small ϕ are the ones most affected by the error. To avoid unrealistic behaviours, in the analysis of the samples with smallest ϕ we introduced the additional rule of setting the slopes of the tielines to be parallel to the closest one.

In this analysis it was necessary to convert c_{PEG} and c_{DNA} from mass density (g/l) to weight ratios. Since we want here to compare our observation with models based on cylinders, we have not used the mass density of dry DNA, $d_{DNA,dry} \approx 1700$ g/l [50], but rather the mass density of hydrated DNA in a cylinder roughly enclosing the duplex, with a diameter of 1.95 nm, which is much larger than the actual volume since it also includes the volume of the two grooves. This assumption leads to a mass density $d_{DNA} \approx 1100$ g/l.

4. Osmotic pressure in DNA LCs

Our results indicate that the phase separation in mixtures of PEG and DNA oligomers is a general phenomenon, which may also involve LC ordering of the DNA-rich phase. LC ordering is however not a necessary condition for phase separation, as demonstrated by liquid–liquid coexistence both in DD-TT solutions and in the DD solutions at low PEG. There are two basic mechanisms that could drive the phase separation. (i) The presence of attractive forces between DNA
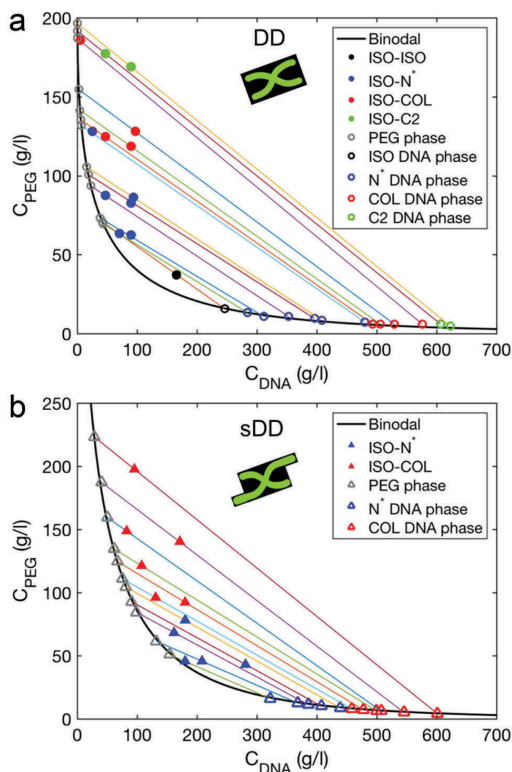
Figure 6. (Colour online) Phase diagrams and phase compositions of DD/PEG-8 kDa/NaCl-300 mM (a) and sDD/PEG-20 kDa/MgCl$_2$-60 mM (b). Full symbols are the initial preparations of a selected subsets of samples. Open symbols express the partitioning of the two components in the two phases: isotropic, PEG-rich phase 1 (grey symbols) and DNA-rich phase 2, which can be in ISO (black symbols), N* (blue symbols), COL (red symbols) and C2 (green symbols) phases. Black lines are binodal lines. Coloured lines are tie lines, connecting the initial preparation to the compositions of the coexisting phases.

duplexes. Such interactions cannot but be between duplex terminal bases. Indeed, in the absence of polycations, side–side interaction between DNA double helices is always repulsive, and can be understood in terms of hydration and electrostatic forces [45]. Base stacking between the blunt ends of DD duplexes, and the mild attraction between the unpaired terminal bases of DD-TT, might in principle justify the phase separation. (ii) Entropic, depletion-type forces due to the difference in flexibility between the two species. The rigid DNA duplexes reduce the conformational space of PEG polymers when they are closer than their gyration radius. Thus, by phase separating, the system reduces the contacts between PEG and DNA,

thus gaining phase space volume [51]. Which of the two mechanisms dominates the DNA-PEG phase separation is at the moment not clear.

At equilibrium, the osmotic pressure of the two phases must be the same, i.e. $\Pi_1 = \Pi_2$. If the partitioning was complete, we could use $c_{PEG,1}$ to determine $\Pi' = \Pi_1(c_{PEG,1}) = \Pi_2(c_{DNA,2})$, where $\Pi_1(c_{PEG,1})$ has been obtained by the already known dependence of PEG osmotic pressure on its concentration [33,34]. Π' values obtained in this way are shown in Figure 7(a) as empty squares. However, the partitioning of PEG and DNA between the two phases is only partial. The concentrations of the minority components in the two phases, despite being smaller than those of the dominant components, could have a relevant effect on the osmotic pressures and their meaning.

Let's first consider the effect of the presence of DNA in phase 1. Since we can safely rule out attractive interactions between PEG and DNA molecules, we can consider DNA as dilute inclusions in a disordered polymer solution. As such, they reduce the free volume of PEG, which is thus effectively at a concentration larger than $C_{PEG,1}$. The free volume subtracted from PEG includes the volume of the DNA duplexes surrounded by a depletion layer having a thickness of the order of the gyration radius of the PEG coils, which for the molecular weight here considered is $R_G \approx 3$ nm (8 kDa) – 5 nm (20 kDa) [52–54]. Since the size of each PEG coil in this study is comparable with the volume of a DNA duplex, we can adopt the simplifying notion that the decrease of the free volume of PEG equals the volume that the DNA duplexes would occupy if they were with no PEG but at the same osmotic pressure: $\Pi_{PEG,1}\left(\frac{c_{PEG,1}}{\phi_{PEG,1}}\right) \approx \Pi_{CS}(f_{DNA,1})$, where we approximate the pressure of the DNA fraction with the CS. $f_{DNA,1}$ is the DNA packing fraction which gives the equilibrium pressure. The volume available to the PEG fraction is $V_{PEG,1} \approx V_{TOT}\left(1 - \frac{c_{DNA,1}}{d_{DNA} f_{DNA,1}}\right)$, so that $\phi_{PEG,1} = V_{PEG,1}/V_{TOT}$. In other words, we are here adopting the notion that both species in the mixture in phase 1 contribute to the osmotic pressure, in line with what observed and predicted for mixtures of hard spheres in the liquid state [55]. The resulting $\Pi'' = \Pi_{PEG,1}\left(\frac{c_{PEG,1}}{\phi_{PEG,1}}\right)$ is plotted in Figure 7(a) as a function of $c_{DNA,2}$ as empty diamond. As visible, the presence of DNA in the PEG-rich phase enhances rather significantly the osmotic pressure at low pressures, when the partitioning is milder, while it becomes negligible at larger pressures, at which the partitioning is virtually complete.

Symmetrically, phase 2, rich in DNA, hosts a minority component of PEG, which also becomes less relevant at larger pressures. At the lowest pressures, phase 2 of DD is in the ISO phase, with $c_{DNA,2} \approx 230$ mg/ml. This corresponds to a density of about one DNA duplexes every 60 nm^3, in turn indicating a mean distance of about 4 nm, clearly compatible with the presence of PEG polymers in the interstitial spaces. Upon increasing the osmotic pressure, the solutions develop cholesteric LC ordering with concentrations $c_{DNA,2}$ in the range 260–420 mg/ml. The mean intercolumnar distance, for this range of concentration, is 3–4 nm, again fully compatible with the intercalation of PEG. At higher pressures, the system turns columnar, and the intercolumnar distance decreases further. This transition is accompanied by a reduction in $C_{PEG,2}$ which becomes virtually negligible.

The effect on the osmotic pressure of PEG in phase 2 can be approximated by considering the fact that, at the interface between phase 1 and phase 2, PEG molecules within the interstitial spaces of DNA in phase 2 are in equilibrium with PEG molecules in phase 1. One could thus assume that volume per molecule of PEG in the interstitial spaces of DNA is equal to that in phase 1, i.e. that the local PEG concentration in both phases is equal to $\frac{c_{PEG,1}}{\phi_{PEG,1}}$, locally providing the same osmotic pressure. Therefore, the osmotic pressure of phase 2 includes the pressure from DNA, which is the quantity we wish to determine, plus the pressure of PEG, which is the same as the one in phase 1 but is exerted only where the interstitial spaces of DNA contact the interface. The surface fraction on which this occurs is given by the fraction of volume filled by PEG in phase 2, i.e. $\frac{c_{PEG,2}\,\phi_{PEG,1}}{c_{PEG,1}}$. We can thus finally determine the osmotic pressure of DNA in phase 2 as the osmotic pressure on the interface between the coexisting phases, reduced by the pressure sustained by the PEG fraction in phase 2: $\Pi_{DNA,2} = \Pi''\left(1 - \frac{c_{PEG,2}\,\phi_{PEG,1}}{c_{PEG,1}}\right)$. These values are also plotted in Figure 7(a) as filled dots. The values are in between the CS equation of state and the perfect gas line.

5. Discussion

In Figure 7(b), we show $\Pi_{DNA,2}$ for the different systems we have studied: DD (dots), sDD (triangles) and DD-TT (squares). Black symbols indicate liquid–liquid phase separations, i.e. phase coexistence in which the DNA-rich phase is in the ISO phase. These are available only for DD at the lowest concentration at which we could find phase separation, and for DD-TT, for
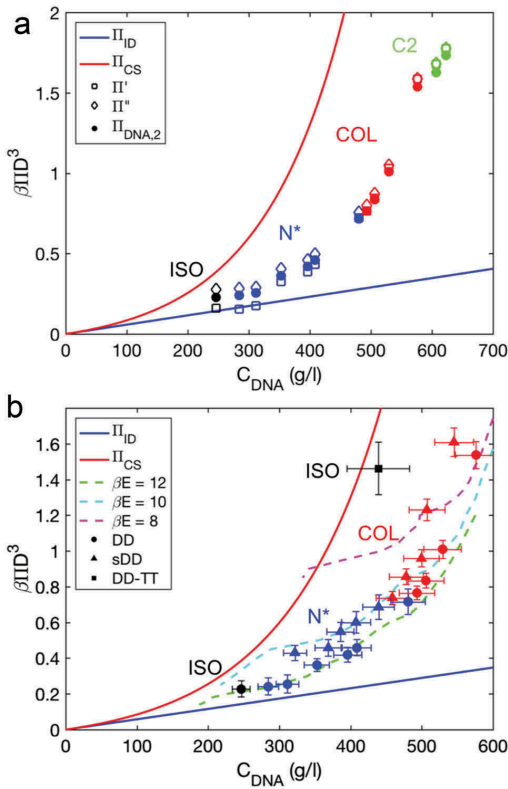
Figure 7. (Colour online) (a) Dimensionless osmotic pressure, $\beta\Pi D^3$, as a function of DNA concentration, c_{DNA}(g/l), measured for DD. Π' (empty squares), Π'' (open diamonds) and $\Pi_{DNA,2}$ (filled circles) indicate the three steps we have adopted for our final estimate of the osmotic pressure in phase 2. Π' is the osmotic pressure of PEG in phase 1. Π'' includes the free volume correction. $\Pi_{DNA,2}$ takes also into account the contribution of the residual PEG in phase 2. (b) Comparison of the experimental equation of state obtained for DD (dots), sDD (triangles) and DD-TT (squares). Error bars are obtained by propagating the experimental uncertainty on the measurement of ϕ. Dashed lines are the prediction of the simulation for a system of sticky cylinders with binding energies as indicated in the legend, from Ref. [11]. (a, b) LC phases are marked as it follows: ISO = isotropic (black), N* = cholesteric (blue), COL = columnar (red) and C2 = crystalline columnar (green). Solid red line is the Carnahan–Starling equation of state for hard spheres and blue line is the ideal gas equation.

which just two experimental points are available (the second located at $\beta\Pi D^3 \approx 3.9$ and $c_{DNA} \approx 570$ mg/ml is not shown because it's out the graph limits) since at lower concentration it does not phase separate. These two points lay closest to the CS equation of state, as expected, since in these cases attractive interactions do not play an important role.

As the LC phases appear, the osmotic pressure vs. concentrations curve move further away from the CS equation of state, as expected because of the larger role played by the attractive interactions and by the increase of free volume due to better packing. In Figure 7(b), we also plot the predictions of the computer simulation of Ref. [11], already shown in Figure 1. Here the predictions have been expressed in terms of DNA concentration by using the density d_{DNA} as discussed above. The data are approximated rather nicely by the coarse-grained model, a condition that enables us to use the simulation to give an estimate of the end-to-end binding energy. As visible, the data behave as the simulations performed with potential depth E in between 10 and 12 $k_B T$, i.e. $E \approx 6$ kcal/mol. Since stacking interactions involve various entropic contributions – because of their hydrophobic nature and because of the conformational stabilization of the terminal bases – the potential E actually represents the association free energy ΔG. In this work, we have explored ΔG only at constant (room) temperature but we expect it to have a large T dependence, as any force involving entropic effect.

Thus, the comparison of our data and the equation of state of Ref. [11]. indicates that, for both DD and sDD, at room temperature $\Delta G \approx 6$ kcal/mol. The similar behaviour of the two systems is fortuitous, since they differ in ionic strength and pH, which are both known to influence the phase separation. Indeed, while sDD has no added salt, DD solutions contain 300 mM NaCl and additional Na$^+$ ions from HEPES buffer. Since salt addition promotes ATPS formation with PEG [23,24] its concentration in phase 2 could in principle be even higher. Added NaCl is known to screen the negative charges of the DNA phosphate backbone and to lower the osmotic pressure of long DNA at low c_{DNA}, a regime where the screened electrostatic is the dominating interaction [45]. For this reason, the pressure obtained for DD could be underestimated in our present work, yielding an overestimation of the stacking ΔG.

The value we find for ΔG is a bit large with respect to literature values but not unreasonable. In the case of sDD duplexes, end-to-end interactions are due to the pairing of the sticky overhangs and can be estimated by using the nearest neighbour model for DNA hybridisation, in analogy to what has been done in Ref. [3]. for the GCCG system. This estimate yields $\Delta G \approx 4$ kcal/mol, i.e. 2 kcal/mol smaller than our evaluation.

The interaction between DD is instead due to stacking forces only. Clean measurements of inter-duplex stacking are still missing. Direct measurements were performed for the interaction between bundles of helices in complex nanostructures [9], obtaining

$\Delta G \approx 3.4$ kcal/mol for each inter-duplex stacking. In this case, however, duplexes don't have the freedom of adjusting their mutual azimuthal angle. Indirect determinations of duplex stacking have been performed by the modelling of LC phase boundaries in solutions of oligomeric duplexes [7] obtaining $\Delta G \approx 3.5$ kcal/mol. In this case, though, the evaluation heavily depends on the use of another coarse-grained model [12]. The stacking interaction between blunt-ended duplexes has been evaluated by atomistic computer simulation by Maffeo et al. [10]. Their result is $\Delta G \approx 6$ kcal/mol, although it is generally believed that stacking is overestimated by this computational approach [56]. While these previous independent estimate of the stacking energy suffer from the limitations mentioned above, the evaluation of the stacking free energy of this work is also affected by experimental uncertainty and by the approximations inherent to the model of Ref. [11]. Among the latter, the most significant involve the assumptions of (i) the cylindrical symmetry of the duplexes and of (ii) the range of interaction. (i) The assumption of cylindrical symmetry implies that the inter-duplex interactions cannot depend on their mutual azimuthal orientations, which is an oversimplification of the interaction potential and of the entropy loss upon binding. By exploring the interaction strength as a function of azimuthal orientation, Maffeo et al. find that the variation in free energy between different choices of the azimuthal angle of two contacting duplexes can be quite consistent, of the order of 2 kcal/mol, indicating that the free adjustment of this parameter, enabled by the free axial rotation of duplexes in the LC phases, might have a large effect on the inter-duplex interactions. At the same time, and because of this very angular dependence of the potential, aggregated duplexes experience a reduced azimuthal freedom, and thus a decreased entropy, which could affect the stacking free energy by approximately 0.5 kcal/mol [12]. (ii) The interaction range adopted in the model of Ref [11] is quite short, corresponding to the expected range of stacking interactions between duplexes in which the terminal bases are bound. However, the base pairs at the terminals may easily melt, thus significantly extending the range in which they can partially base stack. An increased interaction range would imply a smaller positional entropy loss upon binding and thus a smaller depth of the attractive potential.

For these reasons, the gap of ≈ 2 kcal/mol between our finding and the literature values is easily justified by the errors inherent in both our evaluation and the previous ones.

6. Conclusions

We investigated the formation of ATPS in mixtures of PEG and DNA (RNA) duplex-forming oligomers. We found the conditions in which the homogeneous solutions phase separate into coexisting PEG-rich and DNA-rich phases, the latter hosting isotropic, N*, COL or C2 phases. We determined the PEG/DNA phase diagrams for duplexes having different end-to-end interactions – blunt ends (DD and pDD-RNA), sticky overhangs (sDD), aggregation-preventing overhangs (DD-TT) – and different ionic strengths. The resulting binodal lines are well described by the analytical expression empirically obtained for PEG-phosphate ATPS [23].

By measuring the partitioning of the molecular species in the two coexisting phases, we determined the osmotic pressure of DNA duplexes, $\Pi_{DNA,2}$, as a function of the c_{DNA} across the phase diagram, for both isotropic and LC phases. We compared the equation of state obtained in this way with both the CS equation of state for hard spheres and with the pressure predicted by computer simulations of a system of aggregating cylinders [11]. We obtained a good agreement between experiments and simulations, and end-to-end attraction energies of the order of 6 kcal/mol, a bit larger than expected, but still in agreement with the current models for DNA-DNA interactions [10,12].

The present work describes a simple and quick method for measuring the osmotic pressure, Π, and the concentration, c_{DNA}, of DNA LC phases, setting the basis for the usage of DNA/PEG ATPS to finely control the pressures and concentrations of the LC domains. DNA/PEG ATPS could provide an effective strategy to smoothly adjust properties of DNA LCs which are currently hard to control, such as the cholesteric pitch of N*, in turn allowing more accurate measurement of DNA elastic coefficients and non-linear optical response [57,58].

7. Materials and methods

Materials

Lyophilised DNA and RNA oligonucleotides, HPLC purified, are from IDT (DD and pDD-RNA) and from Primm (Milan, Italy) (sDD and DD-TT), while PEG 8 and 20 kDa are from Sigma. Fluorescent PEG 5 kDa is from Nanocs (New York, NY, USA). Stock 50 mg/ml DNA/RNA solutions were obtained by weighting the desired DNA/RNA mass (few milligrams), directly in a 1.5-mL plastic Eppendorf tube, with a Kern (Lohmar, Germany) ABT 100-5M scale,

0.01 mg sensitivity, and adding the required milliQ water volume. Stock PEG solutions, from 0.05 to 0.3 w/w, were prepared by weighting the desired PEG mass (0.5–3 g) and adding the required milliQ water volume in plastic falcon tubes to achieve a 10 g total solution weight.

Samples preparation

Mixtures of DNA and PEG were prepared obtained by depositing controlled volumes (0.5–1 μl) of stock DNA aqueous solution on a microscope glass slide, and letting them evaporate to obtain the desired DNA mass, w_{DNA}. Controlled volumes of PEG solution with mass fraction from 0.05 to 0.30 w/w and with the desired ionic strength and buffers were added then and mixed with a needle. The resulting mixtures were rapidly covered with a second glass slide, sticked together with epoxy glue and sealed with fluorinated oil to ensure stability over time. Cell thickness is set to 20 μm by the use of silica spacers.

Microscope images and analysis

Samples observation and analysis were performed with a Nikon TE-300 microscope allowing bright field, polarised and fluorescence configurations. Images were taken with a Nikon DS-Fi3 colour CCD camera (2880 × 2048 pixels). Single low magnification images (2× objective), and when necessary multiple grid-stitched higher magnification images, were analysed with Fiji-ImageJ software. To obtain the volume fractions of the phases, the area fraction of the fluorescent region and of the total sample area of 8-bit greyscale converted images were measured, after background subtraction, brightness/contrast adjustment and setting of the appropriate threshold.

Authors contributions

T.P.F. and T.B. conceived the experiments; S.D.L., M.T. and T.P.F. performed the experiments; T.P.F. and T.B. analysed the data; and T.P.F., M.T. and T.B. wrote the manuscript.

Acknowledgements

Disclosure statement

The authors declare that they have no competing interests.

References

[1] Nakata M, Zanchetta G, Chapman BD, et al. End-to-end stacking and liquid crystal condensation of 6 to 20 base pair DNA duplexes. Science. 2007;318:1276–1279.

[2] Zanchetta G, Nakata M, Buscaglia M, et al. Liquid crystal ordering of DNA and RNA oligomers with partially overlapping sequences. J Phys Condens Matter. 2008;20:494214.

[3] Fraccia TP, Smith GP, Bethge L, et al. Liquid crystal ordering and isotropic gelation in solutions of four-base-long DNA oligomers. ACS Nano. 2016;10:8508–8516.

[4] Fraccia TP, Smith GP, Clark NA, et al. Liquid crystal ordering of four-base-long DNA oligomers with both G–C and A–T pairing. Crystals. 2018;8. DOI:10.3390/cryst8010005

[5] Salamonczyk M, Zhang J, Portale G, et al. Smectic phase in suspensions of gapped DNA duplexes. Nat Commun. 2016;7:1–9.

[6] Bellini T, Zanchetta G, Fraccia TP, et al. Liquid crystal self-assembly of random- sequence DNA oligomers. Proc Natl Acad Sci U S A. 2012;109:1110–1115.

[7] Smith GP, Fraccia TP, Todisco M, et al. Backbone-free duplex-stacked monomer nucleic acids exhibiting Watson – Crick selectivity. Proc Natl Acad Sci U S A. 2018;115:E7658–E7664.

[8] Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res. 2006;34:564–574.

[9] Kilchherr F, Wachauf C, Pelz B, et al. Single-molecule dissection of stacking forces in DNA. Science (80-.). 2016;353:aaf5508-aaf5508.

[10] Maffeo C, Luan B, Aksimentiev A. End-to-end attraction of duplex DNA. Nucleic Acids Res. 2012;40:3812–3821.

[11] Kuriabova T, Betterton MD, Glaser MA. Linear aggregation and liquid-crystalline order: comparison of Monte Carlo simulation and analytic theory. J Mater Chem. 2010;20:10366.

[12] De Michele C, Rovigatti L, Bellini T, et al. Self-assembly of short DNA duplexes: from a coarse-grained model to experiments through a theoretical link. Soft Matter. 2012;8:8388–8398.

[13] De Michele C, Bellini T, Sciortino F. Self-assembly of bifunctional patchy particles with anisotropic shape into polymers chains: theory, simulations, and experiments. Macromolecules. 2012;45:1090–1106.

[14] Saurabh S, Lansac Y, Jang YH, et al. Understanding the origin of liquid crystal ordering of ultrashort double-stranded DNA. Phys Rev E. 2017;95:1–6.

[15] Carnahan NF, Starling KE. Equation of state for non-attracting rigid spheres. J Chem Phys. 1969;51:635–636.

[16] Horowitz VR, Janowitz LA, Modic AL, et al. Aggregation behavior and chromonic liquid crystal properties of an anionic monoazo dye. Phys Rev E Stat Nonlinear, Soft Matter Phys. 2005;72. DOI:10.1103/PhysRevE.72.041710

[17] Renshaw MP, Day IJNMR. Characterization of the aggregation state of the azo dye sunset yellow in the isotropic phase. J Phys Chem B. 2010;114:10032–10038.

[18] Park HS, Kang SW, Tortora L, et al. Self-assembly of Lyotropic chromonic liquid crystal sunset yellow and

effects of ionic additives. J Phys Chem B. 2008;112:16307–16319.

[19] Kaul A, Pereira RAM, Asenjo JA, et al. Kinetics of phase separation for polyethylene glycol–phosphate two-phase systems. Biotechnol Bioeng. 1995;48:246–256.

[20] Raja S, Murty VR. Liquid-liquid equilibrium of poly (ethylene glycol) 6000 + sodium succinate + water system at different temperatures. The Scientific World Journal 2013;2013:1–7.

[21] Grossmann C, Tintinger R, Zhu J, et al. Aqueous two-phase systems of poly(ethylene glycol) and dextran - experimental results and modeling of thermodynamic properties. Fluid Phase Equilib. 1995;106:111–138.

[22] Zhao Z, Li Q, Ji X, et al. Molar mass fractionation in aqueous two-phase polymer solutions of dextran and poly(ethylene glycol). J Chromatogr. 2016;1452:107–115.

[23] Merchuk JC, Andrews BA, Asenjo JA. Aqueous two-phase systems for protein separation: studies on phase inversion. J Chromatogr B Biomed Sci Appl. 1998;711:285–293.

[24] Asenjo JA, Andrews BA. Aqueous two-phase systems for protein separation: A perspective. J Chromatogr. 2011;1218:8826–8835.

[25] Ferreira AM, Faustino VFM, Mondal D, et al. Improving the extraction and purification of immunoglobulin G by the use of ionic liquids as adjuvants in aqueous biphasic systems. J Biotechnol. 2016;236:166–175.

[26] Monterroso B, Zorrilla S, Sobrinos-Sanguino M, et al. Microenvironments created by liquid-liquid phase transition control the dynamic distribution of bacterial division FtsZ protein. Sci Rep. 2016;6:1–13.

[27] Poudyal RR, Pir Cakmak F, Keating CD, et al. Physical principles and extant biology reveal roles for RNA-containing membraneless compartments in origins of life chemistry. Biochemistry. 2018;57:2509–2519.

[28] Parsegian VA, Rand RP, Rau DC. [3] Macromolecules and water: probing with osmotic stress. Methods Enzymol. 1995;259:43–94.

[29] Annunziata O, Asherie N, Lomakin A, et al. Effect of polyethylene glycol on the liquid-liquid phase transition in aqueous protein solutions. Proc Natl Acad Sci U S A. 2002;99:14165–14170.

[30] Soranno A, Koenig I, Borgia MB, et al. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. Proc Natl Acad Sci. 2014;111:4874–4879.

[31] Pasquier C, Beaufils S, Bouchoux A, et al. Osmotic pressures of lysozyme solutions from gas-like to crystal states. Phys Chem Chem Phys. 2016;18:28458–28465.

[32] Nurmemmedov E, Castelnovo M, Catalano CE, et al. Biophysics of viral infectivity: matching genome length with capsid size. Q Rev Biophys. 2007;40:327–356.

[33] Stanley CB, Strey HH. Measuring osmotic pressure of poly(ethylene glycol) solutions by sedimentation equilibrium ultracentrifugation. Macromolecules. 2003; 36:6888–6893.

[34] Cohen JA, Podgornik R, Hansen PL, et al. A phenomenological one-parameter equation of state for osmotic pressures of PEG and other neutral flexible polymers in good solvents. J Phys Chem B. 2009;113:3709–3714.

[35] Park HS, Kang SW, Tortora L, et al. Condensation of self-assembled lyotropic chromonic liquid crystal sunset yellow in aqueous solutions crowded with polyethylene glycol and doped with salt. Langmuir. 2011;27:4164–4175.

[36] Rau DC, Parsegian VA. Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces. Biophys J. 1992;61:246–259.

[37] Grasso D, Fasone S, La Rosa C. A calorimetric study of the different thermal behaviour of DNA in the isotropic and liquid-crystalline states. Liq Cryst. 1991;9:299–305.

[38] Yasar S, Podgornik R, Valle-Orero J, et al. Continuity of states between the cholesteric → line hexatic transition and the condensation transition in DNA solutions. Sci Rep. 2014;4:1–10.

[39] Zanchetta G, Nakata M, Buscaglia M, et al. Phase separation and liquid crystallization of complementary sequences in mixtures of nanoDNA oligomers. Proc Natl Acad Sci U S A. 2008;105:1111–1117.

[40] Fraccia TP, Smith GP, Zanchetta G, et al. Abiotic ligation of DNA oligomers templated by their liquid crystal ordering. Nat Commun. 2015;6. DOI:10.1038/ncomms7424

[41] Fraccia TP, Zanchetta G, Rimoldi V, et al. Evidence of liquid crystal–assisted abiotic ligation of nucleic acids. Orig Life Evol Biosph. 2015;45:51–68.under subm

[42] Drsata T, Pérez A, Orozco M, et al. Structure, stiffness and substates of the dickerson-drew dodecamer. J Chem Theory Comput. 2013;9:707–721.

[43] Rossi M, Zanchetta G, Klussmann S, et al. Propagation of chirality in mixtures of natural and enantiomeric dna oligomers. Phys Rev Lett. 2013;110:107801.

[44] Zanchetta G, Bellini T, Nakata M, et al. Physical polymerization and liquid crystallization of RNA oligomers. J Am Chem Soc. 2008;130:12864–12865.

[45] Strey H, Parsegian V, Podgornik R. Equation of state for DNA liquid crystals: fluctuation enhanced electrostatic double layer repulsion. Phys Rev Lett. 1997;78:895–898.

[46] Bloomfield V. a DNA condensation by multivalent cations. Biopolymers. 1997;44:269–282.

[47] Wong GCL, Pollack L. Electrostatics of strongly charged biological polymers: ion-mediated interactions and self-organization in nucleic acids and proteins. Annu Rev Phys Chem. 2010;61:171–189.

[48] Raspaud E, Durand D, Livolant F. Interhelical spacing in liquid crystalline spermine and spermidine-DNA precipitates. Biophys J. 2005;88:392–403.

[49] Burak Y, Ariel G, Andelman D. Onset of DNA aggregation in presence of monovalent and multivalent counterions. Biophys J. 2003;85:2100–2110.

[50] Arrighi FE, Mandel M, Bergendahl J, et al. Buoyant densities of DNA of mammals. Biochem Genet. 1970;4:367–376.

[51] Parsegian VA, Rand RP, Rau DC. Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives. Proc Natl Acad Sci. 2000;97:3987–3992.

[52] Linegar KL, Adeniran AE, Kostko AF, et al. Hydrodynamic radius of polyethylene glycol in solution obtained by dynamic light scattering. Colloid J. 2010;72:279–281.

[53] Gurnev PA, Stanley CB, Aksoyoglu MA, et al. Poly (ethylene glycol)s in semidilute regime: radius of gyration in the bulk and partitioning into a nanopore. Macromolecules. 2017;50:2477–2483.

[54] Rubinson KA, Krueger S. Poly(ethylene glycol)s 2000-8000 in water may be planar: A small-angle neutron scattering (SANS) structure study. Polymer (Guildf). 2009;50:4852–4858.

[55] Kranendonk WGT, Frenkel D. Thermodynamic properties of binary hard sphere mixtures. Mol Phys. 1991;72:715–733.

[56] Chen AA, Garcia AE. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. Proc Natl Acad Sci. 2013;110:16820–16825.

[57] Lucchetti L, Fraccia TP, Ciciulla F, et al. Non-linear optical measurement of the twist elastic constant in thermotropic and DNA lyotropic chiral nematics. Sci Rep. 2017;7. DOI:10.1038/s41598-017-05136-z

[58] Lucchetti L, Fraccia TP, Ciciulla F, et al. Giant optical nonlinearity in DNA lyotropic liquid crystals. Opt Express. 2017;25:25951.

Paper 4

Liquid Crystal ordering of DNA Dickerson Dodecamer duplexes with different 5- Phosphate terminations.

I have contributed to this work as follows:

– Experimental design and paper writing;

– 3pDD DNA synthesis and triphosphorylation;

– DD, pDD and 3pDD phase diagram characterization through polarized optical microscopy;

– Cholesteric phase characterization in 3pDD.

Taylor & Francis
Taylor & Francis Group

Check for updates

Liquid Crystal ordering of DNA Dickerson Dodecamer duplexes with different 5'- Phosphate terminations

Marco Todisco[a], Gregory P. Smith[b], and Tommaso P. Fraccia[c]

[a]Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, via Fratelli Cervi 93, I-20090 Segrate (MI), Italy; [b]Department of Physics and Soft Materials Research Center, University of Colorado, Boulder, CO, 80309-0390; [c]Institute Pierre-Gilles de Gennes, ESPCI Paris, PSL Research University, 6 rue Jean Calvin, 75005 Paris, France

ABSTRACT

The onset of liquid crystal (LC) phases in concentrated aqueous solutions of DNA oligomers crucially depends on the end-to-end interaction between the DNA duplexes, which can be provided by the aromatic stacking of the terminal base-pairs or by the pairing of complementary dangling-ends. Here we investigated the LC behavior of three blunt-end 12-base-long DNA duplexes synthesized with hydroxyl, phosphate and triphosphate 5'-termini. We experimentally characterized the concentration-temperature phase diagrams and we quantitatively estimated the end-to-end stacking free energy, by comparing the empirical data with the predictions of coarse-grained linear aggregation models.

The preservation of LC ordering, even in presence of the bulky and highly charged triphosphate group, indicates that attractive stacking interactions are still present and capable of induce linear aggregation of the DNA duplexes. This finding strengthens the potential role of chromonic like self-assembly for the prebiotic formation of linear polymeric nucleic acids.

Introduction

Liquid crystal (LC) ordering of concentrated aqueous solutions of DNA and RNA oligomers has been widely investigated in the recent years [1–12]. The formation of cholesteric, N*, and columnar, COL, LC phases of long nucleic acids, i.e. longer than 100 base pairs (bp), was discovered in the 90s by Livolant and coworkers [13,14], and it is a direct consequence of DNA double helix shape anisotropy and rigidity. Indeed, since DNA persistence length is 50 nm, i.e. of the order of 150 bp [15], polymeric DNA duplexes are surely hard rod-like molecules, which satisfy the Onsager's constraint for liquid crystallization [16]. In the oligomeric regime (<20 bp), which is characterized by helical fragments whose length, $L < 6.8$ nm, is comparable to the diameter, $D = 2$ nm, the onset of LC phases takes place as consequence of the formation of physical reversible linear aggregates, overcoming the otherwise unsatisfied Onsager's rule of $L/D > 4$ [16]. Here, the key factor is the end-to-end interaction of short DNA duplexes, 6 to 20 bp, which can happen through the stacking of the terminal bases of blunt-end duplexes [1,2] or by pairing of complementary terminal dangling ends [5], which

therefore become "sticky-ends". LC ordering has been extended to DNA sequences whose length doesn't permit double helix stability at room temperature, such as DNA 4mers [9,10]. Here the key factor is the simultaneous pairing and chaining of nicked strands which stabilizes the whole aggregate's self-assembly. Even more surprising is the recently reported discovery that single bases DNA and RNA nucleotides (dNTPs and rNTPs) can give rise to COL phase when brought to high concentration and low temperature [4]. Being such monomeric level the minimal nucleic acids based system capable of pairing and stacking, it indicates that the formation of LC phases is ubiquitous property of nucleic acids, regardless of the presence of the sugar-phosphate backbone [17]. Despite the several hydrogen bonding motifs which are accessible to canonical nucleobases, this happens only when the Watson-Crick pairing rule is satisfied, i.e. only if A-T or G-C pairing is possible [4]. In this latter case the nucleotides have 5' - triphosphate group, which plays the essential role of enhancing monomers solubility, even if such effect is still to be investigated more deeply.

The richness and robustness of oligo DNA and RNA LC phase phenomenology recently inspired the idea that it could have played a role for the prebiotic formation of polymeric nucleic acids [18]. Indeed, the self-assembly steps involved in this phenomenon can be seen as a process to enhance the formation of linear molecular products, which takes its foundation from the shape and interaction asymmetry of the involved building blocks. Which is lastly a propagation of the symmetry breaking derived by the properties of hydrogen bonding and stacking of aromatic nucleobases. In this frame, the onset of LC ordering of DNA and RNA oligomers, between 6 and 14 base-pairs long, is capable to boost the rate of non-enzymatic ligation reaction, increasing both reaction yield and product length, and favoring linear products against circular ones [19–21]. This process which has been named "LC autocatalysis" could have played a crucial role in series of events that brought to the origin of life on our planet [22].

The ligation of DNA or RNA strands requires the presence of a phosphate group at 5'- (or 3'-) termini, which must bind to a 3'- (or 5'-) hydroxyl termini of another strand. Normally this reaction does not happen spontaneously, since the opposite process, the phosphodiester bond hydrolysis, is favored in aqueous environments, unless the ligation rate is enhanced by the coupling of some molecular moieties which work as good leaving groups. Water soluble carbodiimides, as EDC or CDI, are the most efficient phosphate activating agents [23,24]. Imidazole compounds are good leaving groups too, with a reasonable prebiotic availability [25]. The breaking of a triphosphate group, with the production of pyrophosphate as leaving group, is the strategy adopted by the modern enzymes, such as ligases or polymerases. Despite triphosphate ligation is fairly effective in absence of the catalyzing enzymes or a templating strand [26,27], it could have provided a simple and straightforward prebiotic ligation strategy. Given the crucial role of phosphate termini for chemical reactions and for lack of a deep understanding of the way different phosphate termini perturbate the stacking of DNA duplexes, we carried a systematic study of the effect of three different 5' termini on the formation of liquid crystal phases. In particular, we investigated the LC behavior of the already well characterized Dickerson Dodecamer [1,4,28–30], in the presence of three different 5'- molecular terminations: 1) 5'- OH group, i.e. the standard DD duplex with no terminal phosphates; 2) 5'– single phosphate group, i.e. the 5'- phosphorylated DD duplex,

named pDD; 3) 5'– tri-phosphate group, i.e. the 5'- tri-phosphorylated DD duplex, named 3pDD. We experimentally characterized the concentration, c_{DNA}, vs. temperature, T, phase diagram of these three systems and we quantitatively estimated the end-to-end stacking free energy, by comparing empirical data with the predictions of coarse-grained linear aggregation models in which DNA helices are simplified as sticky cylinders [31,32].

Synthesis and purification

While the synthesis of DNA/RNA oligomers with OH termini or with 3' or 5' single phosphate modification is standard nowadays and can be easily obtained from the major nucleic acids producing companies, the synthesis of 5'-triphosphate oligomers is not so usual and commercially available. Therefore, we decided to synthesize by ourselves all the DNA sequences involved in this work. The synthesis of 5'-hydroxil Dickerson Dodecamer (DD, 5'-CGCGAATTCGCG-3'), 5'-phosphate (pDD, 5'-pCGCGAATTCGCG-3') and 5'-triphosphate (3pDD, 5'-pppCGCGAATTCGCG-3'), were executed using an Äkta Oligopilot by Amersham Pharmacia Biotech. Standard solid support approach was followed to obtain both DD and pDD. Triphosphorylation on the 5'-terminus of DD was achieved following a modified protocol from [33]. Oligos were phosphitylated directly on the synthesizer using one column volume of 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one (1M solution of 1:4 dioxane to pyridine) followed by reaction with tributyammonium pyrophosphate (0.5M solution of 1:1 acetonitrile to pyridine). Both mixtures were protected by air oxidation working under Argon flux. The reactions were injected directly into the column and recycled for $\sim$ 20 minutes to reach reaction completion. Column content was recovered, and DNA molecules were detached from the solid support by stirring for 8 hours the in 40ml of NH$_4$OH solution (28 $\sim$ 30% w/w) in a sand bath at 50 °C. In order to get rid of the beads, the suspension was cooled down to room temperature and vacuum filtered over a coarse frit.

The filtered solution was stripped of solvent under vacuum using a Heidolph rotary evaporator. Final volume was reduced to $\sim$10ml. RP-HPLC was performed through a prep scale SepaxGP-C8 column (Sepax Technologies Inc) on a dual Varian Prostar 218 system tied to a Prostar 318 detector to yield purified 3pDD. The analyte was eluted with a linear gradient using 50 mM TEAA pH 7.0 as the primary buffer and Methanol as the secondary buffer.

The RP-HPLC elution fractions were characterized through MALDI experiments on a AB/Sciex Voyager DE-STR instrument. Samples from the collected fractions were serially diluted and spotted on the MALDI plate using 3-hydroxypicolinic acid as a matrix. We could confirm that 3pDD was present in the first eluted peak, with a retention time between 30 and 32 minutes. Analysis of chromatograph reveals an efficiency for the tri-phosphorilation reaction close to 5%.

3pDD fractions were pooled and salt-precipitated adding a NaCl (final concentration equal to 400mM) and an approximately equal volume of isopropanol. Precipitation was performed with centrifugation at 15'000g and 25 °C for 10 minutes. Pellet was resuspended in 2ml 70% ethanol solution. The resulting solution was centrifuged at 15'000g
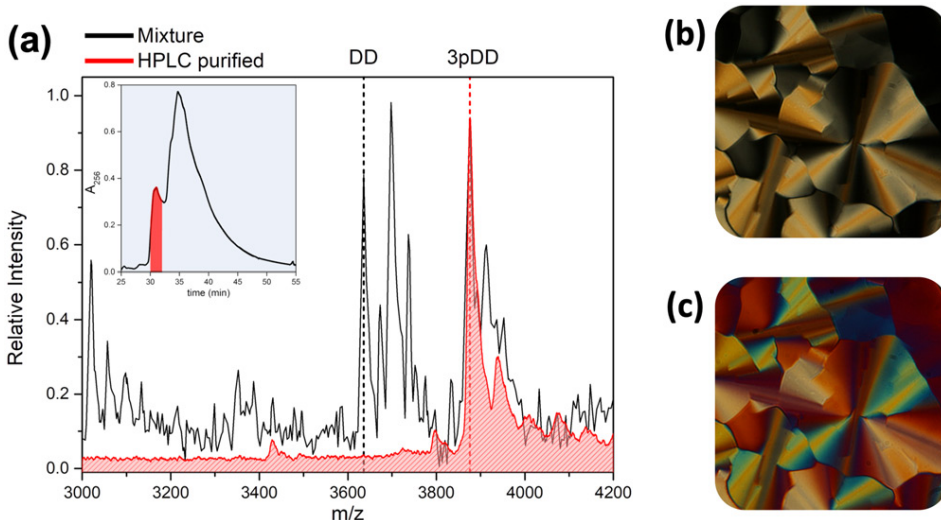
Figure 1. **Synthesis and purification of 3pDD.** (a) MALDI-TOF spectra of unpurified 3pDD synthesis product (black line) and after HPLC purification (red line). Elution peak corresponding to 3pDD is highlighted in red in the chromatograph (inlet). (b, c) Purified 3pDD readily produces liquid crystalline phases at high concentration, showing columnar focal conic textures observable through polarized transmitted optical microscopy (b) and with the addition of quarter-wave plate compensator (c). The combined observation of micrographs b and c allow to estimate a negative Δn for 3pDD, which is the same as for pDD and DD DNA oligos. This suggesting that the same ordering mechanism is at the basis of LC phases appearance.

and 25 °C for 10 minutes. The pellet was air-dried at 37 °C and the remaining pellet was resuspended in 2 ml of milliQ water and lyophilized.

Results and discussion

Characterization of 3pDD liquid crystal phases

To investigate the formation of LC phases for 3pDD duplexes, we produced glass planar microscope cells on which 1 µl drops were deposited, from a stock solution prepared at $c_{DNA} = 50$ mg/ml, and then let to dry. When LC phases appear, the cells were covered with a second glass, glued and sealed with fluorinated oil, Fomblin Y25 Solvay, to prevent evaporation. Cell thickness of 20 µm between glass plates was guaranteed by silica rods spacers.

Fig. 1b,c and Fig. 2 show that 3pDD exhibits typical N* and COL textures. In particular, the analysis of columnar focal conics with a quarter-wave plate compensator, see Fig. 1b,c, shows that Δn is negative, as in the case for all DNA LC phases. The observation of the cholesteric textures reveals that the pitch, i.e. the periodicity of the twisted cholesteric planes along the direction perpendicular to the planes themselves, is in the micrometer range, Fig. 2a. The measurement of the pitch in the range between 0 °C and 10 °C shows that the pitch slightly increases with temperature, see graph in Fig. 2a. This behavior is the same previously observed for DD N*, while surprisingly pDD was reported to have a pitch decreasing with T [7]. It has to be taken into account that more recently also DD pitch has been reported to decrease with temperature [29]. The
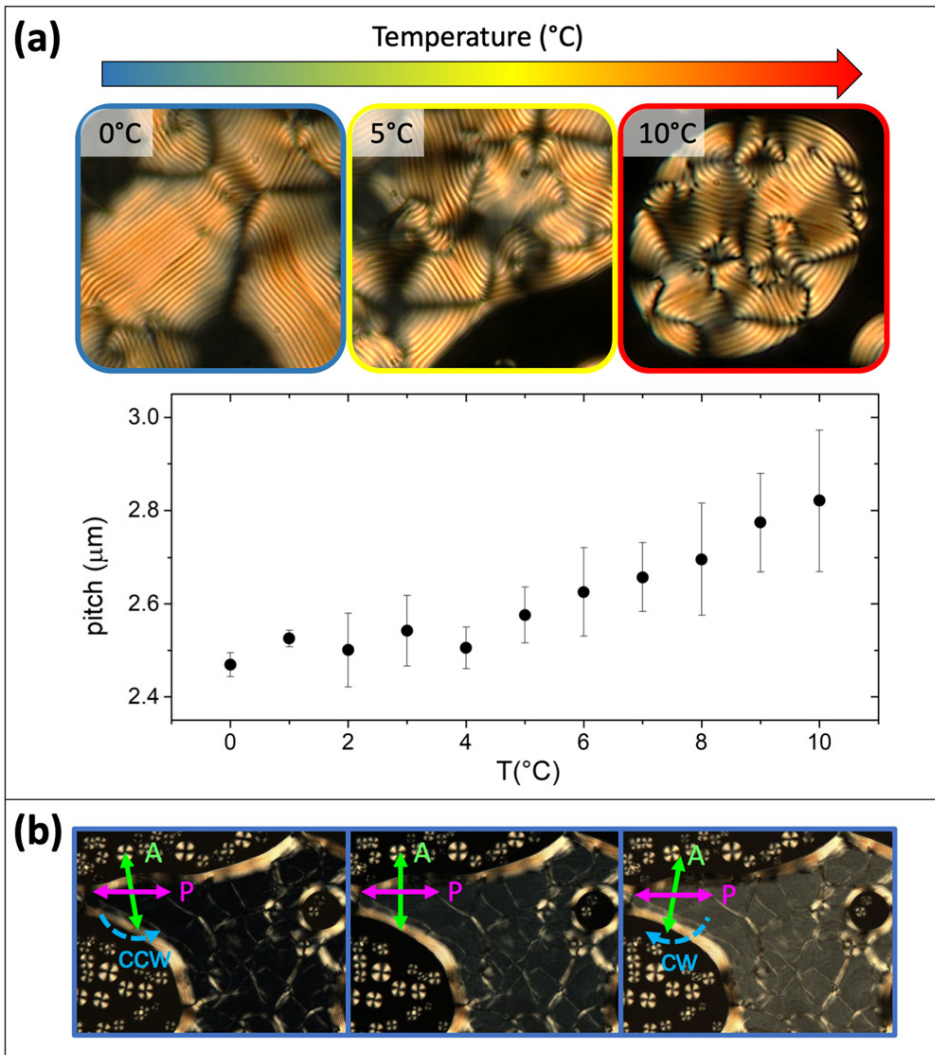
Figure 2. Characterization of cholesteric LC phases in 3pDD. (a,b) 3pDD can self-organize in cholesteric LC in a narrow range of concentrations ($c_{3pDD} \approx 450$ - $500\,mg/ml$) with a pitch in the micrometer range. (a) The analysis of the periodicity of the characteristic fingerprints shows a positive dependence on temperature. (b) Transmitted light through cholesteric textures can be extinguished by rotating the analyzer counter-clockwise, indicating that the N* phase of 3pDD is left-handed.

disagreement between different measurements of the pitch for the same molecule can be explained by the different purification grade of the DNA oligomers, which can cause huge differences of ionic strength, prevalently due to Na^+ ions, and pH [9]. Both ionic strength and pH can tune the electrostatic repulsion between the charged backbone of the DNA duplexes, having effect on the final twist between neighboring molecules.

The handedness of the cholesteric phase can be determined by de-crossing the polarizers. As shown in Fig. 2b, the cholesteric domains having axes perpendicular to the cell plates, i.e. along the direction of observation, are bluish when observed between crossed
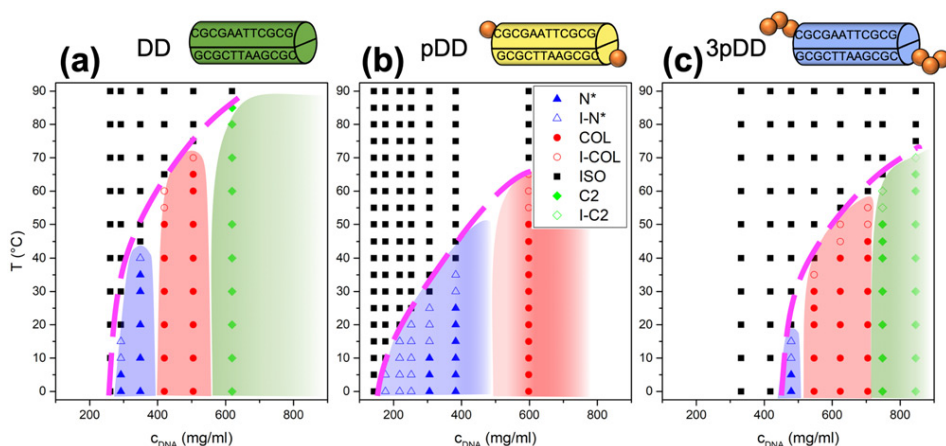
Figure 3. Supramolecular liquid crystal ordering in aqueous solutions of DD oligomers with different 5'-terminus functional groups. (a, b, c) Concentration, c_{DNA}, vs. temperature, T, phase diagrams of DD (a), pDD (b) and 3pDD (c) show that the presence of tri-phosphate groups at the 5'-terminus reduces the stability of the LC phases and increases the concentration required to undergo the liquid crystalline transition.

polarizers, see central micrograph of Fig. 2a. The bluish color is due to selective reflection of the N* phase having the pitch in the infrared. The cholesteric domains can be extinguished only when the analyzer is de-crossed counter-clockwise, left micrograph, and not when it's rotated clockwise, right micrograph. This indicates that the N* phase of 3pDD is left-handed. This is quite surprising since both DD and pDD have been previously reported to display right-handed cholesterics [7]. This apparent contradiction can be explained once more by the different conditions in which the experiments were carried. Right-handed N* were observed in samples with ISO-N* transition, at 25 °C, that was taking place at very high concentrations, $c_{DNA} = 730$ mg/ml for DD and $c_{DNA} = 670$ mg/ml for pDD in ref. [7] and $c_{DNA} = 610$ mg/ml for DD in ref. [29]. These values are more than the double of those observed here, see phase diagrams in Fig. 3, and this difference is almost surely due to the difference in the ionic strength and pH of the DNA solutions. The DNA sequences used here, after the HPLC purification, were dissolved in milliQ water at $c_{DNA} = 5$ mg/ml, and were dialyzed overnight in 1L of milliQ water and then in 1L of NaCl 50 mM, in order to achieve the same ionic strength condition.

The handedness of the N* phase is ruled by the balance between the steric interaction between DNA duplexes, which favors the propagation of the right-handedness of the natural enantiomer helix, and the electrostatic repulsion, which instead favors the opposite left-handed chirality [34,35]. The balance of these two interactions depends on the distance between DNA duplexes, and therefore on DNA concentration, the first prevailing at high c_{DNA} and the latter at low c_{DNA}. In this frame it's not surprising that the N* phase of 3pDD could be left-handed, since it is found at $c_{DNA} \asymp 480$ mg/ml.

Comparison of the phase diagrams of DD, pDD and 3pDD

The formation LC phases indicated that, even when the dodecamers are triphosphate terminated, there are attractive stacking interactions between the terminal nucleobases

of blunt-ended duplexes. To assess the presence and strength of stacking interaction for DNA duplexes with the different 5'-phospate termination, we compared the phase behavior of aqueous solution of the dodecamers DD, pDD and 3pDD, which are sketched in the upper part of Fig. 3. Glass capillaries were also prepared by collecting 1 to 2 mg of lyophilized DNA powder, as measured with a 0.01 mg sensitivity scale, ABT 100-5M Kern, within a borosilicate capillary 0.5 mm diameter with a single pre-sealed end. By adding the desired volume of water with an Eppendorf P2.5 micropipette from the open end of the capillary and measuring the weight increase, we obtained a highly accurate method to investigate the phase diagram. The c_{DNA} vs. T phase diagrams for DD, pDD and 3pDD are shown in Fig. 3a,b,c, where the black, blue, red and green symbols indicate the ISO, N*, COL and higher order columnar, C2, phases, respectively. Open symbols indicate the coexistence between isotropic and the LC phase, relative to the same color code. The comparison between the phase boundaries, reported in Fig. 3, clearly indicate that stacking interactions are weaker in 3pDD, since LC phases appear only at larger c and lower T, with respect to both DD and pDD. This appears reasonable given the larger electrostatic repulsion and of the bulkier terminal group provided by triphosphates.

To quantitatively investigate the observed phase behavior, we used a coarse-grained model in which DNA duplexes are treated as hard cylinders with rounded edges that associate because of a square-well attractive interaction [31,32]. The terminal edges are decorated with a sticky semi sphere whose dimension indicates the range of the attractive potential. This model enables the calculation of the phase boundaries from geometrical and interaction parameters. In the isotropic phase, the cylinders aggregate forming chains of M monomers. The average aggregation number $\langle M \rangle$ depends on the ratio between the inter-monomer binding strength ΔG and the thermal energy $k_B T$ as

$$\langle M \rangle = \frac{1}{2} \left(1 + \sqrt{1 + 8\phi \frac{v_B}{v_D} e^{\Delta G / k_B T}} \right)$$

where φ is the volume fraction filled by cylinders, i.e. rescaled DNA concentration; v_B is the "bond volume", i.e. the volume that the center of mass of each cylinder can explore when part of an aggregate; v_D is the volume of a cylinder. According to the model the phase boundary marking the ISO side of ISO-LC coexistence is found when

$$\langle M \rangle \approx 0.75 \ \phi^{-1.76} D/L$$

with D and L being the diameter and length of the cylinders, respectively. In the case of DD, pDD and 3pDD, $D/L \asymp 0.5$, since D = 2 nm and L = 12·0.34 nm = 4.08 nm. The volume fraction of DNA duplexes, φ, is obtained by assuming DNA density as 1687 g/l, as reported in ref. [4].

The stacking free energy ΔG between terminals of distinct bundles of DNA duplexes was evaluated experimentally [36]. The value obtained for CG/GC contacts of the kind involved in DD stacking, is about -2.15 kcal/mol, a value close to the one typically used for the same CG/GC contact in the nearest neighbor model for thermodynamic stability of DNA [37]. Another indirect evaluation of the stacking free energy ΔG for DD duplexes in LC phase, has been also obtained by exploiting the phase separation between the flexible polymer PEG and oligo DNA duplexes [38]. In this case,

the comparison of the pressure, P, vs. c_{DNA} phase diagram, with the prediction of Montecarlo simulations from ref. [39], led to the estimation of $\Delta G \asymp 6\,\text{kcal/mol}$. Since the measurement in Ref. [36] does not provide the enthalpic, ΔH, and entropic, ΔS, contributions to ΔG separately, we decided to adopt those of the broadly adopted model of Ref. [37]: $\Delta H = -9.8\,\text{kcal/mol}$ and $\Delta S = -24.4\,\text{cal/(mol K)}$, or $\Delta G \sim 2\,\text{kcal/mol}$.

These values, plus the geometrical parameters of the DD duplexes, are not sufficient to calculate the phase diagram. To compute $\langle M \rangle$ it is necessary to know the ratio v_B/v_D, a value which is very difficult to estimate without a detailed knowledge of the stacking potential. Moreover, having adopted in the calculation nearest neighbor values, which are obtained in the context of the internal interaction in chemically continuous double helices, the bond volume must also effectively include a correction for the entropy associated to the azimuthal freedom of the interacting duplexes.

To compare the values of ΔG for the stacking of duplexes with added phosphates with the value of the OH terminated duplexes, we varied both the interaction strength and the parameter v_B/v_D, to approximate the phase boundaries. Thus, we uniformly changed both enthalpy and entropy by a multiplicative factor A, which gives $\Delta H' = A \cdot \Delta H$ and $\Delta S' = A \cdot \Delta S$ and the parameter v_B/v_D in order to approximate the phase boundaries. The experimental ISO-LC phase boundaries and their approximation from the coarse-grained model are reported in Fig. 4, as filled data points and straight lines, respectively.

When $A = 1.55$ and $v_B/v_D = 0.3$ the phase boundary best matches the one of DD (black line in Fig. 4). To evaluate how much the stacking energy should be reduced to shift the DD phase boundary toward that of 3pDD, we proceeded via two different routes. First, we kept $v_B/v_D = 0.3$ fixed and reduced the energy. When $A = 0.8$ the phase boundary best matches the one of 3pDD (green line in Fig. 4). Second, since the presence of a terminal phosphate has been shown to stabilize the azimuthal freedom of the stacked duplexes [40], we held $\Delta G'$ as in DD and decreased the value of v_B/v_D. We find that when $v_B/v_D \asymp 0.02$, the predicted phase boundary approximates that of 3pDD (green light line in Fig. 4). Finally, to match the phase boundaries of pDD, we kept the reduced v_B/v_D as found in 3pDD, and we uniformly changed enthalpy and entropy. When $A = 2.2$ the phase boundary best matches the one of pDD (red line in Fig. 4).

In order to compare the obtained stacking energies, we incorporated the reduced bond volume into the free energy value yielding: $\Delta G_{DD} = 2.73\,\text{kcal/mol}$, $\Delta G_{pDD} = 2.74\,\text{kcal/mol}$, $\Delta G_{3pDD} = 1.05\,\text{kcal/mol}$ for $v_B/v_D = 0.3$, and $\Delta G_{3pDD} = 1.07\,\text{kcal/mol}$ for the reduced $v_B/v_D = 0.02$, at $37\,^\circ\text{C}$.

Thus, this analysis suggests that the free energy of stacking of triphosphate terminated duplexes is much lower of the one involved in non-phosphorylated duplexes, around 2.5 times. This can be explained by the reasonable entropic penalty given by the steric occupation of the triphosphate chains and the high electrostatic repulsion of such charged moieties. On the other hand, the addition of a single phosphate isn't affecting appreciably the interaction strength. This is in agreement with the results of atomistic simulations which show that the phosphate terminus enhances the azimuthal lockage of the interacting duplexes [40].
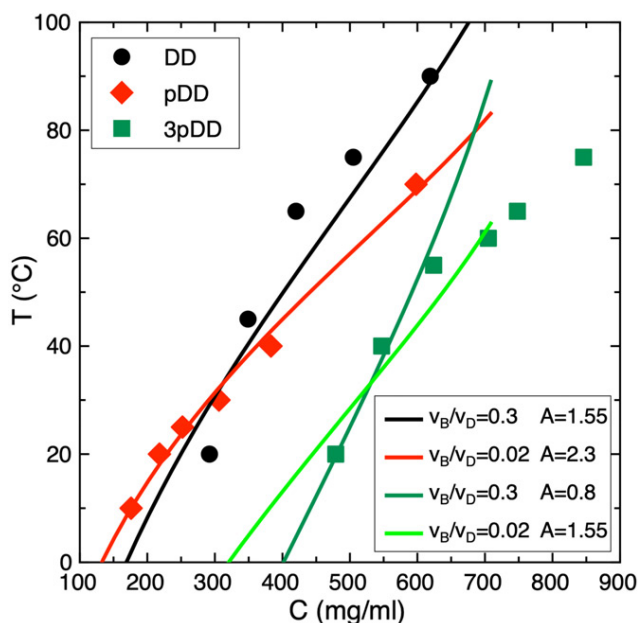
Figure 4. Comparison of the phase boundaries of DD, pDD and 3pDD with the prediction of the coarse-grained model. The experimental c_{DNA} and T data of the ISO-LC phase boundaries for DD, pDD and 3pDD are reported by the filled symbols, in black circles, red diamonds and green squares, respectively. The lines are the predicted phase boundaries from the coarse-grained model of ref. [31], that better approximate the experimental data. The simulation parameters v_B/v_D and A, which are reported in the legend, allow the calculation of the intermolecular end-to-end stacking free energy at 37 °C for the three investigated DNA duplexes: ΔG_{DD} = 2.73 kcal/mol, ΔG_{pDD} = 2.74 kcal/mol, ΔG_{3pDD} = 1.05 kcal/mol for v_B/v_D = 0.3, and ΔG_{3pDD} = 1.07 kcal/mol for v_B/v_D = 0.02.

Conclusions

The LC phase diagrams of three version of the 12mer DNA sequence, 5'-CGCGAATTCGCG-3', with 5'-OH, 5'-P and 5'-PPP termini, have been investigated and quantitatively compared. Despite the potentially destabilizing, variety of chemical modification of the 5'-termini, LC ordering is preserved in all cases. LC ordering of nucleic acids stabilizes the physical continuity of long double helixes formed by the aggregation of short duplexes despite the different chemical terminations. The persistence of LC phases in triphosphate terminated DNA duplexes strengthens the potential role of nucleic acids self-assembly in the context of the origin of life [18,22]. Thus, the reported results indicates that the mechanism of LC autocatalysis, that we previously suggested [19–21], is applicable even in the presence of bulky, charged and, importantly, potentially reactive terminal groups for the abiotic polymerization of nucleic acids.

Acknowledgement

M.T., G.P.S. and T.P.F. conceived the experiments; M.T. and G.P.S. performed the DNA synthesis; M.T., G.P.S. and T.P.F. performed the experiments; M.T. and T.P.F. analyzed the data; T.P.F. wrote the manuscript.

The authors declare that they have no competing interests.

References

[1] Nakata, M., Zanchetta, G., Chapman, B. D., Jones, C. D., Cross, J. O., Pindak, R., Bellini, T., & Clark, N. A. End-to-end stacking and liquid crystal condensation of 6 to 20 base pair DNA duplexes. *Science* 2007, *318*, 1276–1279, doi:10.1126/science.1143826.

[2] Zanchetta, G., Bellini, T., Nakata, M., & Clark, N. A. Physical polymerization and liquid crystallization of RNA oligomers. *J. Am. Chem. Soc.* 2008, *130*, 12864–12865, doi:10.1021/ja804718c.

[3] Lucchetti, L., Fraccia, T. P., Ciciulla, F., & Bellini, T. Non-linear optical measurement of the twist elastic constant in thermotropic and DNA lyotropic chiral nematics. *Sci. Rep.* 2017, 7, doi:10.1038/s41598-017-05136-z.

[4] Smith, G. P., Fraccia, T. P., Todisco, M., Zanchetta, G., Zhu, C., Hayden, E., Bellini, T., & Clark, N. A. Backbone-free duplex-stacked monomer nucleic acids exhibiting Watson–Crick selectivity. *Proc. Natl. Acad. Sci.* 2018, 201721369, doi:10.1073/pnas.1721369115.

[5] Zanchetta, G., Nakata, M., Buscaglia, M., Clark, N. A., & Bellini, T. Liquid crystal ordering of DNA and RNA oligomers with partially overlapping sequences. *J. Phys. Condens. Matter* 2008, *20*, 494214, doi:10.1088/0953-8984/20/49/494214.

[6] Zanchetta, G., Nakata, M., Buscaglia, M., Bellini, T., & Clark, N. A. Phase separation and liquid crystallization of complementary sequences in mixtures of nanoDNA oligomers. *Proc. Natl. Acad. Sci. U. S. A.* 2008, *105*, 1111–1117.

[7] Zanchetta, G., Giavazzi, F., Nakata, M., Buscaglia, M., Cerbino, R., Clark, N. a, & Bellini, T. 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.* 2010, *107*, 17497–17502, doi:10.1073/pnas.1011199107.

[8] Bellini, T., Zanchetta, G., Fraccia, T. P., Cerbino, R., Tsai, E., & Smith, G. P. Liquid crystal self-assembly of random- sequence DNA oligomers. *Proc. Natl. Acad. Sci. U. S. A.* 2012, *109*, 1110–1115, doi:10.1073/pnas.1117463109.

[9] Fraccia, T. P., Smith, G. P., Bethge, L., Zanchetta, G., Nava, G., Klussmann, S., Clark, N. A., & Bellini, T. Liquid Crystal Ordering and Isotropic Gelation in Solutions of Four-Base-Long DNA Oligomers. *ACS Nano* 2016, *10*, 8508–8516, doi:10.1021/acsnano.6b03622.

[10] Fraccia, T. P., Smith, G. P., Clark, N. A., & Bellini, T. Liquid crystal ordering of four-base-long DNA oligomers with both G–C and A–T pairing. *Crystals* 2018, 8, doi:10.3390/cryst8010005.

[11] Di Leo, S., Todisco, M., Bellini, T., & Fraccia, T. P. Phase separations, liquid crystal ordering and molecular partitioning in mixtures of PEG and DNA oligomers. *Liq. Cryst.* 2018, 1–13, doi:10.1080/02678292.2018.1519123.

[12] Lucchetti, L., Fraccia, T. P., Ciciulla, F., Simoni, F., & Bellini, T. Giant optical nonlinearity in DNA lyotropic liquid crystals. *Opt. Express* 2017, 25, doi:10.1364/OE.25.025951.

[13] Livolant, F., Levelut, A. M., Doucet, J., & Benoit, J. P. The highly concentrated liquid-crystalline phase of DNA is columnar hexagonal. *Nature* 1989, *339*, 724–726.

[14] Leforestier, a, & Livolant, F. Supramolecular ordering of DNA in the cholesteric liquid crystalline phase: an ultrastructural study. *Biophys. J.* 1993, *65*, 56–72, doi:10.1016/S0006-3495(93)81063-4.

[15] Lu, Y., Weers, B., & Stellwagen, N. C. DNA persistence length revisited. *Biopolymers* 2001, *61*, 261–275, doi:10.1002/bip.10151.

[16] Onsager, L. The effects of shape on the interaction of colloidal particles. *Ann. N. Y. Acad. Sci.* 1949, *51*, 627–659.

[17] Podgornik, R. Sticking and stacking: Persistent ordering of fragmented DNA analogs. *Proc. Natl. Acad. Sci.* 2018, *115*, 8652–8654, doi:10.1073/pnas.1810662115.

[18] Budin, I., & Szostak, J. W. Expanding roles for diverse physical phenomena during the origin of life. *Annu. Rev. Biophys.* 2010, *39*, 245–63, doi:10.1146/annurev.biophys. 050708.133753.

[19] Fraccia, T. P., Smith, G. P., Zanchetta, G., Paraboschi, E., Yi, Y., Yi, Y., Walba, D. M., Dieci, G., Clark, N. A., & Bellini, T. Abiotic ligation of DNA oligomers templated by their liquid crystal ordering. *Nat. Commun.* 2015, *6*, 6424, doi:10.1038/ncomms7424.

[20] Fraccia, T. P., Zanchetta, G., Rimoldi, V., Clark, N. A., & Bellini, T. Evidence of Liquid Crystal–Assisted Abiotic Ligation of Nucleic Acids. *Orig. Life Evol. Biosph.* 2015, 45, doi: 10.1007/s11084-015-9438-1.

[21] Todisco, M., Fraccia, T. P., Smith, G. P., Corno, A., Bethge, L., Klussmann, S., Paraboschi, E. M., Asselta, R., Colombo, D., Zanchetta, G., Clark, N. A., & Bellini, T. Nonenzymatic Polymerization into Long Linear RNA Templated by Liquid Crystal Self-Assembly. *ACS Nano* 2018, acsnano.8b05821, doi:10.1021/acsnano.8b05821.

[22] Lazcano, A. Prebiotic Evolution and Self-Assembly of Nucleic Acids. *ACS Nano* 2018, acs-nano.8b07605, doi:10.1021/acsnano.8b07605.

[23] Dolinnaya, N. G., Tsytovich, a V, Sergeev, V. N., Oretskaya, T. S., & Shabarova, Z. a Structural and kinetic aspects of chemical reactions in DNA duplexes. Information on DNA local structure obtained from chemical ligation data. *Nucleic Acids Res.* 1991, *19*, 3073–3080.

[24] Shabarova, Z. A., Merenkova, I. N., Oretskaya, T. S., Sokolova, N. I., Skripkin, E. A., Alexeyeva, E. V, Balakin, A. G., & Bogdanov, A. A. Chemical ligation of DNA: 1991, *19*, 4247–4251.

[25] Schrum, J. P., Ricardo, A., Krishnamurthy, M., Blain, J. C., & Szostak, J. W. Efficient and rapid template-directed nucleic acid copying using 2'-amino-2',3'-dideoxyribonucleoside-5'- phosphorimidazolide monomers. *J. Am. Chem. Soc.* 2009, *131*, 14560–14570.

[26] Rohatgi, R., Bartel, D. P., & Szostak, J. W. Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. 1996, *9*, 3332–3339.

[27] Rohatgi, R., Bartel, D. P., & Szostak, J. W. Nonenzymatic, Template-Directed Ligation of Oligoribonucleotides Is Highly Regioselective for the Formation of 3 ′ - 5 ′ Phosphodiester Bonds. 1996, 3340–3344, doi:10.1021/ja9537134.

[28] Drsata, T., Pérez, A., Orozco, M., Morozov, A. V., Soner, J., & Lankas, F. Structure, stiffness and substates of the dickerson-drew dodecamer. *J. Chem. Theory Comput.* 2013, *9*, 707–721.

[29] De Michele, C., Zanchetta, G., Bellini, T., Frezza, E., & Ferrarini, A. Hierarchical Propagation of Chirality through Reversible Polymerization: The Cholesteric Phase of DNA Oligomers. *ACS Macro Lett.* 2016, *5*, 208–212, doi:10.1021/acsmacrolett.5b00579.

[30] Rossi, M., Zanchetta, G., Klussmann, S., Clark, N. a., & Bellini, T. Propagation of chirality in mixtures of natural and enantiomeric dna oligomers. *Phys. Rev. Lett.* 2013, *110*, 107801, doi:10.1103/PhysRevLett.110.107801.

[31] De Michele, C., Bellini, T., & Sciortino, F. Self-assembly of bifunctional patchy particles with anisotropic shape into polymers chains: Theory, simulations, and experiments. *Macromolecules* 2012, *45*, 1090–1106.

[32] Nguyen, K. T., Sciortino, F., & De Michele, C. Self-assembly-driven nematization. *Langmuir* 2014, *30*, 4814–4819, doi:10.1021/la500127n.

[33] Lebedev, A. V, Koukhareva, I. I., Beck, T., & Vaghefi, M. M. Preparation of oligodeoxynu-cleotide 5′-triphosphates using solid support approach. *Nucleosides, Nucleotides and Nucleic Acids* 2001, *20*, 1403–1409, doi:10.1081/NCN-100002565.

[34] Frezza, E., Ferrarini, A., Kolli, H. B., Giacometti, A., & Cinacchi, G. The isotropic-to-nem-atic phase transition in hard helices: Theory and simulation. *J. Chem. Phys.* 2013, *138*, 24, doi:10.1063/1.4802005.

[35] Kolli, H. B., Frezza, E., Cinacchi, G., Ferrarini, A., Giacometti, A., & Hudson, T. S. Communication: From rods to helices: Evidence of a screw-like nematic phase. *J. Chem. Phys.* 2014, *140*, 10–15, doi:10.1063/1.4866808.

[36] Kilchherr, F., Wachauf, C., Pelz, B., Rief, M., Zacharias, M., & Dietz, H. Single-molecule dissection of stacking forces in DNA. *Science (80-.)* . 2016, *353*.

[37] SantaLucia, J., & Hicks, D. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 2004, *33*, 415–40, doi:10.1146/annurev.biophys.32.110601.141800.

[38] Di Leo, S., Todisco, M., Bellini, T., & Fraccia, T. P. Phase separations, liquid crystal ordering and molecular partitioning in mixtures of PEG and DNA oligomers. *Liq. Cryst.* 2018, *00*, 1–13, doi:10.1080/02678292.2018.1519123.

[39] Kuriabova, T., Betterton, M. D., & Glaser, M. a. Linear aggregation and liquid-crystalline order: comparison of Monte Carlo simulation and analytic theory. *J. Mater. Chem.* 2010, *20*, 10366, doi:10.1039/c0jm02355h.

[40] Maffeo, C., Luan, B., & Aksimentiev, A. End-to-end attraction of duplex DNA. *Nucleic Acids Res.* 2012, *40*, 3812–3821.