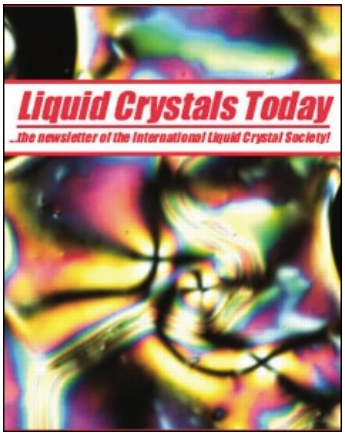

Liquid Crystals Today
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713681230

Spontaneous self-assembly of nucleic acids: liquid crystal condensation of complementary sequences in mixtures of DNA and RNA oligomers

Giuliano Zanchetta[a]
[a] Dipartimento di chimica, biochimica e biotecnologie per la medicina, Universita' degli studi di Milano, Italy

To cite this Article Zanchetta, Giuliano(2009) 'Spontaneous self-assembly of nucleic acids: liquid crystal condensation of complementary sequences in mixtures of DNA and RNA oligomers', Liquid Crystals Today, 18: 2, 40 — 49
To link to this Article: DOI: 10.1080/13583140903155093
URL: http://dx.doi.org/10.1080/13583140903155093

Liquid Crystals Today
Vol. 18, No. 2, December 2009, 40–49

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Spontaneous self-assembly of nucleic acids: liquid crystal condensation of complementary sequences in mixtures of DNA and RNA oligomers

Giuliano Zanchetta*

Dipartimento di chimica, biochimica e biotecnologie per la medicina, Universita' degli studi di Milano, Italy

(Received 20 January 2009; final form 31 March 2009)

"Le cose tutte quante
hanno ordine tra loro, e questo e' forma
che l'universo a Dio fa simigliante."
(Dante Alighieri, *Divina Commedia, Paradiso*, I, 103)

[All things whate'er they be
Have order among themselves, and this is form,
That makes the universe resemble God]

A remarkable example of spontaneous self-assembly of biomolecules, driven purely by physical interactions, is reported. Short, complementary DNA and RNA oligomers, down to six bases in length, exhibit lyotropic liquid crystal phases, chiral nematic and columnar, although these duplexes lack the shape anisotropy required for liquid crystal ordering. Such phases are produced by the end-to-end stacking of the duplex oligomers into polydisperse, anisotropic, rod-shaped aggregates, which can order into liquid crystals. Furthermore, when only a fraction of the sample is composed of complementary sequences, and hence the solution is effectively a mixture of single strands and double-stranded helices, the system is found to phase separate with duplex-rich liquid crystalline domains emerging from an isotropic background rich in single strands. This spontaneous partitioning, resulting from a combination of entropic and enthalpic factors, is sensitive to the degree of complementarity of the sequences and can be tuned with temperature. We suggest that in a chemical environment where oligomer ligation is possible, such ordering and condensation could provide a plausible route for the selective synthesis of extended complementary oligomers, a mechanism of possible relevance in prebiotic scenarios.

Keywords: DNA; RNA; nucleic acids; stacking; chromonics; phase transitions; Onsager; nucleation; depletion; prebiotic; ribozyme; ligation

1. Introduction

One of the most challenging and fascinating goals of science is to account for the wonderful variety and complexity of the biological world in terms of simple and general laws. Towards this aim, a key step is to refer to notions from statistical and soft matter physics, whose application to biological mechanisms has provided remarkable successes. Examples range from protein folding (*1*), to virus formation (*2*), to cellular organisation, whose complexity can be reduced to the relative simplicity of self-assembled structures, driven by packing constraints and enforced by van der Waals, hydrophobic and electrostatic interactions (*3*,*4*). An exciting challenge is thus to find the intrinsic self-assembly properties driving the spontaneous ordering of biological macromolecules, also in the frame of their emergence from prebiotic chemistry (*5*), and to control and mimic natural assembly for materials design.

The formation of reversible or irreversible aggregates with elongated shapes is a distinctive property of a variety of molecules, such as the stacks of polyaromatic dyes and drug molecules (*6*), stiff bundles of cytoskeleton proteins (*4*) and amyloid fibers (*7*). Perhaps the most studied process occurs in surfactant solutions, with aggregates growing in one, two or three dimensions to form rod-like, plate-like or spherical micelles (*8*).

In 1949, Onsager showed theoretically that the hard-core excluded volume interaction alone is sufficient to induce orientational alignment of elongated particles in solution, other anisotropic interactions being not necessary (*9*); in this way, the onset of the nematic liquid crystalline (LC) phase can be explained. Since then, entropy has often been involved as a source of order. In this vein, by combining colloidal hard rods and hard spheres, a phase diagram is obtained that is much richer than those occurring in separate systems of spheres or rods, with a variety of layered and columnar arrangements (*10*). Spontaneous, entropy-driven mechanisms such as orientational and positional order and demixing, well characterised in colloids, liquid crystals and polymers, are being increasingly recognised as also playing a central role in many biological systems too (*3*, *4*).

*Email: giuliano.zanchetta@unimi.it

ISSN 1358-314X print/ISSN 1464-5181 online

DOI: 10.1080/13583140903155093
http://www.informaworld.com

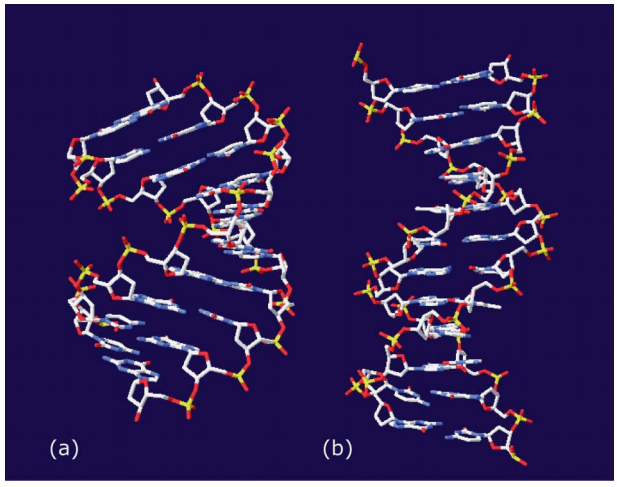

Figure 1. The most common structures displayed by nucleic acids. (a) Crystallographic structure of double-stranded DNA oligomer CCGGGCCCGG in the A form (PDB code 1ZF1); (b) DNA oligomer CGCGAATTCGCG in the B form (PDB code 1BNA).

In this respect, DNA (shown in Figure 1 in its most typical A and B helical forms) is an especially interesting example. Its packing *in vivo* is very tight, with high volume fractions in bacteria and in virus capsids, and it displays an astonishing hierarchical assembly into chromosomes, compacting metres of double helix down to a few micrometres (*11*). The packing mechanisms adopted are strongly related to its natural tendency, as a long, semi-rigid polymer, to form lyotropic liquid crystalline phases in concentrated aqueous solutions, as observed *in vitro* (*12*). Since the early studies (*13*, *14*), two main mesophases have been recognised, chiral nematic and hexagonal columnar (*15–17*), but blue and hexatic phases have also been reported (*17*, *18*). The LC phases of solutions of duplex B-form DNA have been extensively characterised by optical (*17*), X-ray (*19*), and magnetic resonance (*20*) methods for chain lengths from 10^6 down to 10^2 base pairs (bp), comparable in size to the B-DNA bending persistence length, $L_p \sim 50$ nm (150 bp). A semi-quantitative agreement with Onsager theory (*21*) and computer simulations (*22*) is found, and better results are obtained if finite length, flexibility and electrostatic interactions are taken into account (*21*, *23*, *24*).

In addition to being a valuable model for appraising LC theories, the ordered packing of DNA may have important physiological roles, as a natural strategy to protect the genome of various bacteria under prolonged environmental stress, such as starvation and intense radiation, reducing the exposition to damage factors and templating DNA repair (*25*, *26*).

In short, DNA acts as a liquid crystal because of its elongated shape, and nature exploits such a tendency to pack and preserve it. However, how did it happen, in the prebiotic era, that DNA and the more ancient RNA *acquired* their polymeric structure without the cell machinery? How was the free energy barrier to elongation overcome? And how did nucleotides emerge as the building blocks for the carriers of genetic information, the major prerequisite to life?

Some new light is shed on these outstanding questions by the experiments presented here, on the LC behaviour of ultrashort DNA and RNA duplexes, down to 6 bp (with an aspect ratio of about one), which have never been observed before and was unexpected because such short duplexes are weakly anisotropic in shape.

The origin of our research (a joint project between Professor Tommaso Bellini in Milan and Professor Noel Clark and Dr Michi Nakata in Boulder, CO) was the intention to use nucleic acids as building blocks to construct molecules with tuneable length and rigidity. In fact, by exploiting the well-known pairing mechanism of DNA and the relevant difference in flexibility between paired and unpaired sequences, macromolecular assemblies can be created with various sizes and mechanical properties. Along this line, two sequences were chosen first: CGCAATTGCG, which is self-complementary, i.e. can hybridise with another strand of the same sequence and thus form a rigid, squat helix; and CGCAATTGCGTTTTTTTTTT, which is only partially self-complementary, forming a helix with two long, dangling tails. This latter structure, a rigid core with flexible ends, bearing resemblance with usual liquid crystal molecules, should have favoured the formation of LC phases, whereas the rigid helix without tails was the negative control, because its aspect ratio, lower than two, should prevent spontaneous alignment. What we found was just the opposite of what we expected: the oligonucleotide with tails showed no mesophases, while the 'control' oligo displayed wonderfully coloured textures when observed in a polarised microscope, just like much longer DNA strands. Soon, understanding the origin of the LC phases in such short DNA duplexes became the main interest. We then characterised the phase diagram of different sequences with various lengths and duplex terminals, finding evidence that the stacking interactions between neighbouring nitrogen bases, which ultimately determine the helical shape of DNA (*11*), promote end-to-end aggregation of helices and the formation of longer chains, capable of alignment (*27*). Indeed, in the partially self-complementary strand the presence of dangling tails disturbs the terminal adhesion and thus suppresses the liquid crystals.

We also studied the effect of the denaturation of the helices on the liquid crystals: when roughly half of the helices are broken by heat (and thus single-stranded DNA (SS-DNA) coils appear), the phases also vanish.

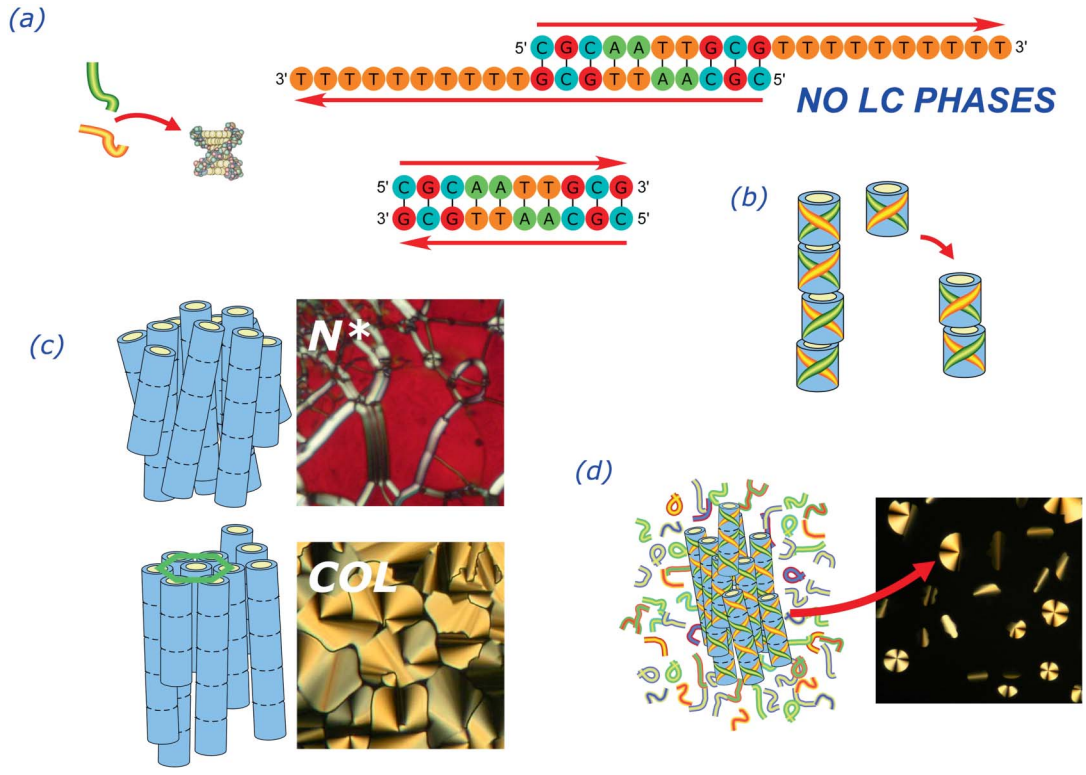

Figure 2. Sketch of the self-assembly process leading to the LC alignment of well-paired short DNA duplexes. (a) Complementary and partially complementary sequences pair into helices. While duplexes with overhangs do not show LC behaviour, blunt-end duplexes can be idealised as hydrophilic cylinders with hydrophobic ends capable of end-to-end adhesion and stacking (b) into units sufficiently anisotropic to orientationally and positionally order into LC phases (c), chiral nematic N* and hexagonal columnar COL. (d) Upon cooling a mixture of single-stranded nDNAs, the complementary oligomers (yellow–orange and yellow–green) form duplexes and phase separate into LC domains.

At this point, the second misapprehension came to light: to obtain a system intrinsically composed of helices and coils at equilibrium, we mixed mutually complementary sequences, CCTCAAAACTCC and GGAGTTTTGAGG, in unbalanced ratios. At some SS-DNA concentration, LC phases were expected to become disturbed and disappear. Surprisingly, they did not: a phase separation occurred, instead, between the helices and the pool of coils, with double-stranded DNA (DS-DNA) and thus complementary sequences organising in liquid crystal droplets (*27*,*28*). The same behaviour occurs for RNA (*29*), and the whole process is sketched in Figures 2: (a) pairing, (b) end-to-end adhesion, (c) ordering and (d) demixing combine to yield LC domains of helices.

We suggest that this staged self-assembly may have been instrumental in early life as a means of templating the linear polymer structure of information carriers, since in the presence of appropriate chemistry, end-to-end aggregation and phase separation would strongly promote the elongation of complementary oligomers. To substantiate this hypothesis, in the following sections we describe some relevant features of DNA and RNA LC phases and the main factors controlling their nucleation in mixtures of sequences.

2. Nano-DNA liquid crystals

The appearance of LC phases in long DNA (lDNA, $N > 100$ bp) has been accounted for theoretically by modelling B-DNA as a repulsive rigid or semi-flexible rod-shaped solute (*9*, *21*). The complete computer-simulated phase diagram for hard rods by Bolhuis and Frenkel (*22*) quantitatively confirms this picture for length-to-diameter ratios $L/D > 4.7$ ($N > 28$ bp), and also shows that for $L/D < 4.7$ there should be no LC phases at any volume fraction ϕ. Nevertheless, surprisingly, we found nematic ordering for ultrashort B-DNA duplexes (nDNA) of length 6 bp $< N <$ 20 bp (2–7 nm in length). At the lower end of this scale the L/D ratio is about one and the LC phase is appearing for N values an order of magnitude smaller than those predicted by the Onsager–Bolhuis–Frenkel (OBF) criterion, indicating that interactions in addition to excluded volume effects must be present.

We studied a series of self-complementary nDNA, duplex-forming oligomers, including the well-studied

'Dickerson dodecamer' (D12) CGCGAATTCGCG, along with a large variety of non-complementary and partially complementary oligomers. nDNA solutions in thin glass cells were investigated through a wide range of techniques: polarised transmission optical microscopy (PTOM) to investigate optical textures; fluorescence optical microscopy to probe the orientation of DNA strands and duplex formation; optical reflection interferometry to measure the refractive index and concentration c (in milligrams solute per millilitre solution) of solutions; and synchrotron microbeam X-ray diffraction (XRD) to probe local molecular organisation (for details, see (*27*)).

In spite of the challenges presented by the extremely small nDNA sample quantities available (elegantly overcome also thanks to Michi Nakata's gradient cells providing the whole sequence of phases within a narrow channel through controlled evaporation), these techniques provided unambiguous evidence for chiral nematic (N*) and uniaxial, hexagonal columnar (COL) liquid crystal phases in the nDNA solutions. Owing to the underlying aggregation process, such phases can be regarded as direct analogues of the chromonic phases N* and M (see (*6*)). At higher concentration, higher-order columnar (C_2) and crystal-like (X) phases were also found, whose structure is currently under investigation.

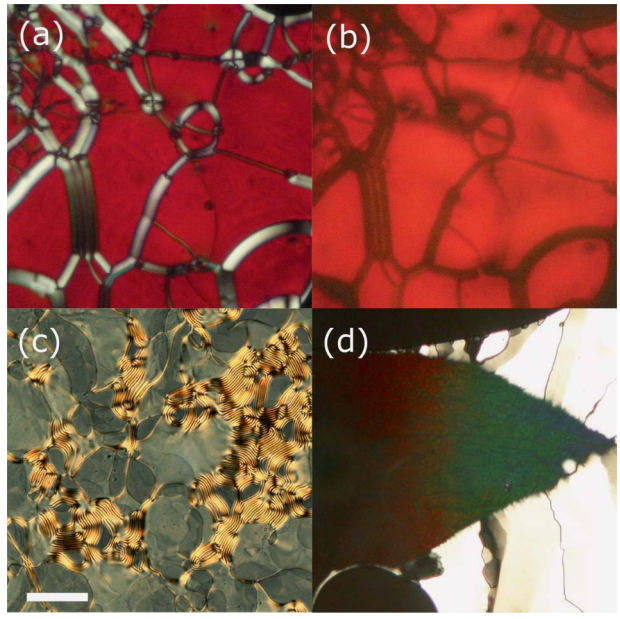

Figure 3. Optical microscopy textures of the chiral nematic phase of nDNA and nRNA sequences. (a) PTOM image of oily streaks in a 12 bp nDNA sample. (b) Unpolarised reflected light image of the same sample, showing red selective reflection, corresponding to a cholesteric pitch of ~450 nm. (c) Oily streaks in a nDNA 12 bp sample with 'fingerprints' in the defect regions, revealing a pitch well above 1 -μm. (d) 12 bp nRNA: the pitch decreases with increasing concentration, from left to right. Scale bar is 20 -μm in (a) and (b) and 50 -μm in (c) and (d).

2.1 *Cholesteric phase N**

Since DNA molecules are chiral, their nematic phase exhibits macroscopic chirality in the form of a helical precession in space of the local optical axis (the local DNA molecular axis; note that the DNA helix has negative birefringence). The N* phase appears with its chiral helix axis z either parallel (N_{PAR}) or normal (N_{NOR}) to the plates, forming either the Grandjean (sometimes focal conic, if the cholesteric pitch p is smaller than 1 -μm) or fingerprint (if $p > 1$ -μm) textures in the PAR case, and forming Grandjean terraces in the NOR case. The oily streaks textures, such as those shown in Figure 3(a) and (c), combine the two features.

The cholesteric pitch was observed to increase with nDNA concentration c and with temperature T, approaching the N*–I transition. This dependence corresponds to the behaviour of lDNA (*17, 30, 31*) and the substantial continuity between nDNA and lDNA LC phases was verified through a 'contact' cell between a 10 bp sequence and a ~ 900 bp sample (*27*).

However, as is clear from the selective reflection colours in the visible range exhibited by the N* phase of nDNA in unpolarised reflection microscopy (Figure 3(b)), the pitch was found for several of the oligomers to be as short as $p \sim 350$ nm, considerably smaller than the 2–4 -μm typically observed in $N \geq 147$ bp DNA and explained by the current models (*32*, *33*). Actually, there are also some incidental observations (*34*) of sub-micrometre pitches in lDNA phases, while on the other hand we occasionally observe fingerprints also in the shortest sequences, because of the non-negligible influence of the boundary conditions. However, there is a clear positive trend in the average pitch, depending on the length of the oligos and no sub-micrometre p was ever detected for $N \geq 14$. We are currently exploring the connections between the chiral behaviour and the flexibility of the duplex chains.

2.2 *Hexagonal columnar phase COL*

A higher concentration phase grows from the isotropic phase upon cooling, mainly with developable domains, as seen in Figure 4(a). Such textures are clear indications of either fluid lamellar smectic or fluid columnar order with respectively the layer normal or column axis parallel to the glass plane. A study of the developable domain birefringence with a variable wave-plate compensator showed that the high refractive index direction is radial where the domains are circular, indicating that

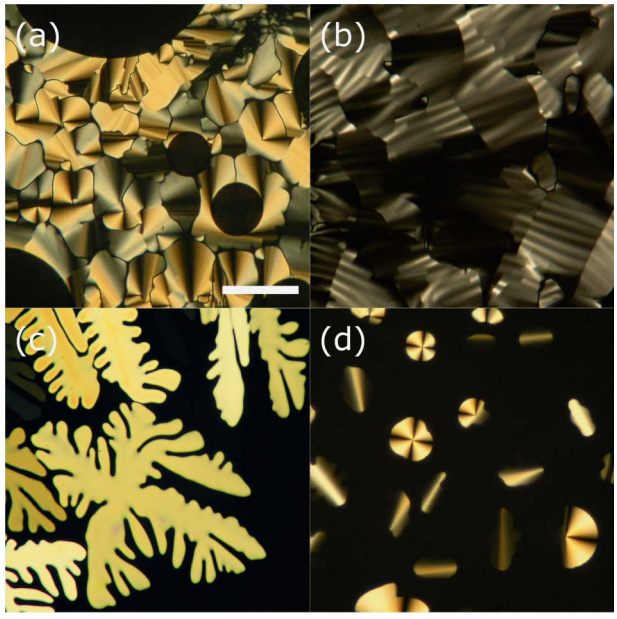

Figure 4. PTOM textures of the hexagonal columnar phase of some nDNA samples. (a) Developable domains in a DNA 12 bp sample. (b) Undulating texture in an equimolar mixture of the complementary sequences CCTCAAAACTCC (A) and GGAGTTTTGAGG (B). (c) Dendritic structures of a 12 bp sample cooled from the isotropic phase. (d) COL domains surrounded by isotropic phase in a mixture of the sequences A and B with molar ratio [B]/[A] ~ 3. Scale bar is 100 -μm in (a)–(c) and 40 -μm in (d).

the DNA base pair planes are also radial and thus that the DNA double helix axes lie parallel to circles about the domain centre. Following the arguments of Livolant formulated for lDNA (*16*), the structure is therefore columnar (COL), with the double helix axes parallel to the columns, and in the birefringent domains the column axis is parallel to the glass plane (COL_{PAR}). The molecular arrangement of the COL phase can also be simply displayed by doping the nDNA with ethidium bromide (EB), a fluorescent molecule prone to intercalation between paired DNA bases (*35*). Since its fluorescence absorption and emission are strongly anisotropic, higher in the plane of the molecule, the dependence of fluorescent emission on the polarisation of incident light gives an indication of intercalated EB orientation, and thus on the orientation of the host helices. An unpolarised fluorescence image gives a 'map' of the z arrangement of the molecules (not shown), while by changing the polarisation direction of the incident light, we can obtain unambiguous confirmation on the tangential arrangement of helices within the developable domains. For homeotropically aligned helices, the EB molecules, being perpendicular to the incident light, have the same fluorescent emission for every incident linear polarisation; for planar helices, instead, the intensity of fluorescence depends on the direction of helices on the plane, as shown in Figures 5(c) and (d).

As regards the nature of the columnar phase, it was investigated by XRD; owing to the small amount of material available, and the poor control of the boundary conditions, which prevents us from obtaining large aligned samples, we performed simultaneous PTOM and synchrotron-based microbeam XRD to collect diffraction patterns from small ($\sim 14 \times 14$ -μm^2), selected areas: evidence for hexagonal packing in the plane perpendicular to the helices was found (*27*).

Overall, the optical and diffraction experiments indicate unambiguously that the COL phase is a uniaxial fluid columnar liquid crystal, in line with the earlier findings for lDNA. This is also confirmed by measuring the specific birefringence for the COL phase at the N* boundary ($\Delta n \sim 0.025$ at 350 mg ml^{-1}), which yields values in close agreement with those measured for the lDNA LC phases, once properly rescaled for concentration, thus supporting the notion that the packing of the nDNA phase is strictly analogous to that of the lDNA LC phases. Finally, this conclusion is also supported by a contact preparation, wherein the COL phase is also continuous along the 10 – 900 bp concentration gradient (*27*), and by other textures observed in PTOM: undulating textures (Figure 4(b)), already reported for the COL phase of lDNA (*17*), and dendrites growing upon cooling from the isotropic phase (Figure 4(c)), also found in discotic liquid crystals.

2.3 *Evidence for end-to-end interactions*

The DNA concentration was calculated from the refractive index n, which in turn was measured through microscope-based interferometry (*27*): the observed Fabry–Perot fringes of light reflected from the glass/solution interfaces enabled extraction of the optical path through the solution, and thus the local value of n. LC phases are found in the sub-Onsager region of the phase diagram of nDNA where they are not expected on the basis of duplex shape. Bearing in mind that the basic features of LC ordering are common for nDNA and lDNA and that nematic and columnar LC ordering can appear in weakly anisotropic solutes by equilibrium end-to-end self-assembly into linear chains of particles (*23*), we are lead to the conclusion that the origin of the LC phases in nDNA is the physical 'living polymerisation' of short duplexes into extended duplex units that are long and rigid enough to order. A similar hypothesis was proposed by Alam and Drobny in attempting to account for an orientationally ordered phase in nuclear magnetic resonance (NMR) studies of a

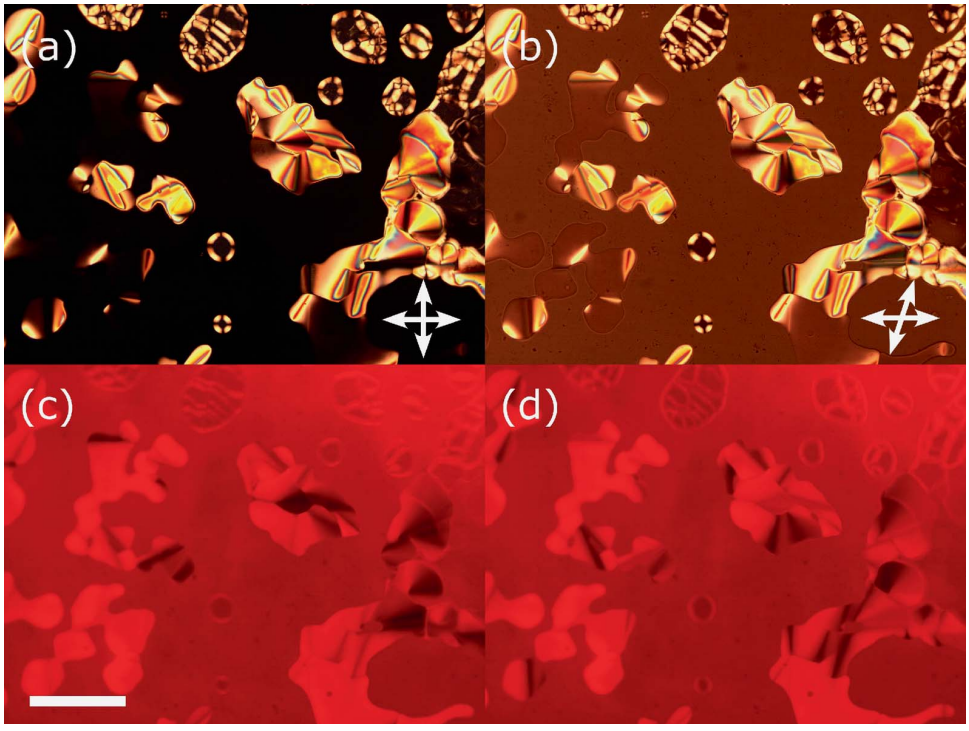

Figure 5. Arrangement of nDNA helices within the COL phase. (a) PTOM image of columnar domains of a 12 bp nDNA sample doped with EB. Some LC areas are black because of the homeotropic alignment of duplexes, as revealed by slightly de-crossing the analyser (b). By imaging the same sample with fluorescence microscopy, the planar regions have strongly anisotropic signal as the incident light has vertical (c) or horizontal polarisation (d), depending on the orientation of the helices inside the domains; the homeotropic regions, instead, have fluorescent emission independent of incident polarisation. Scale bar is 100 -μm.

B-DNA dodecamer (to the best of the author's knowledge, the only previous report of LC phases displayed by oligonucleotides) (*20*). Indeed, end-to-end adhesion suppresses the lamellar smectic A phase predicted for monodisperse hard rods by favouring end-to-end rather than side-by-side positional correlations and by introducing unavoidable length polydispersity into the ordering units, both effects reducing the entropic free-energy gain of lamellar ordering, in agreement with our observations. This sort of assembly is familiar in crystalline nDNA (*36*) and in nDNA/protein complexes (*37*, *38*), as well as in lyotropic chromonic liquid crystals (*6*, *39*), and has been studied extensively for particular DNA bases. For example, Guanosine forms H-bonded tetramers aggregating in a chromonic-like process into stacks, which in turn can form mesophases (*40*). Whilst blunt end stacking is of little importance when dealing with long DNA, it becomes the major feature controlling the self-assembly and phase behaviour of short DNA. Indeed, Mundy *et al.* (41) considered the chromonic stack as an effective model for the dynamic state of the central column of bases in DNA.

Computer simulations (*42–44*) show that for sufficiently rigid aggregate chains, the I–N transition occurs according to the OBF prediction if the average aggregate length $<L>$ is used in the OBF model; such agreement thus enables us to extract, from the nDNA phase diagram, an estimate of the stacking free energy associated with the reversible aggregation of duplexes, $\Delta E \sim 4$–$8\, k_B T$ (see (*27*)). Such values are similar to the base stacking energy found for G-quartets (*40*) and for the combination of pairing and stacking within the DNA helix (*45*); similar values also characterise chromonic aggregation (*39*).

In an effort to influence the duplex end-to-end adhesion we carried out further experiments on the D12, but with added unpaired tails, with 1T, 2T or 10T groups added at the 3' terminal. We found that this modification suppresses the LC phases with the exception of the C_2 phase, which is still observed in the case of D12 with 1T and 2T groups. We interpret this result as indicating that dangling ends impede end-to-end adhesion, in line with the base-stacking concept. In contrast, D12 duplexes phosphorylated at the 5′ end exhibited nearly the same phase behaviour as that of the –OH terminated D12 duplexes described so far. Such duplexes, bearing one phosphate group per end, have the same composition as obtained by fracturing a

lDNA duplex; in contrast, the addition of one more charged phosphate group on terminals (that is, phosphorylation of duplexes at both 3′ and 5′ terminals) again suppressed the liquid crystals, as with the 1T and 2T terminations. Finally, a different effect of the terminals modification is obtained through a sequence such as CGCGCGAATTCGCG, the same as D12 but bearing an additional CG at the 5′ end. This sequence, when forming a duplex, has dangling ends capable of forming links with neighbouring duplexes. For this reason, such a sequence is not disfavoured, but is actually forced to form chains somewhat analogous to nDNA living polymers. Indeed, such an oligomer shows similar phase behaviour as D12, but with lower critical concentrations and higher thermal stability, a fact consistent with the higher energy associated with the CG pairing and stacking compared with the blunt end stacking (*46*). This observation could provide a new route for the estimation of the free energy associated in the Watson–Crick pairing of two DNA or RNA strands into double helices, by changing the sequence of the 'sticky' dangling tails.

3. Nano-RNA

From a structural point of view, RNA shares some basic features with DNA, but it usually adopts an A-conformation (Figure 1(a)) with terminal bases tilted $\sim 20^\circ$ with respect to the normal to the helical axis, while in the B-form the tilt is $\sim 0^\circ$. However, concentrated solutions of self-complementary RNA oligomers exhibit similar self-assembly into N* and COL phases (Figure 3(d)), indicating that this newly observed form of spontaneous ordering is universal in nucleic acids (*29*). The tilted base structure corresponds to a crucial difference between DNA and RNA since the straightness of the aggregate required for LC ordering implies for RNA the matching of the tilt of the terminals, in turn constraining the relative positions of the sugar–phosphate chains at their chemical discontinuities along the weakly aggregated physical polymer. This constraint appears particularly favourable for promoting ligation at the gaps of the phosphate chain, should the chemical environment be suitable for catalysing such a reaction.

Overall, the cholesteric pitch p is shorter in nRNA than in nDNA for any particular oligo length, probably reflecting the shorter helical pitch of the A-form structure. Remarkably, it also displays an opposite dependence on c. Experiments on long DNA chains have revealed a non-monotonic dependence of p on concentration (*31*), and theoretical results indicate that p critically depends on the spectrum of the helical charge distribution with the potential for p versus c dependence of either signs (*33*). Hence, the origin of RNA peculiarity could just reside in the differences between the A- and B-helix forms and in the azimuthal continuity of the phosphate chains along the aggregated physical polymers.

4. Helices and coils

The nDNA solutions exhibit thermotropic mesomorphism, melting at sufficiently high T, to the optically isotropic liquid (I) phase. One of the key factors controlling the presence of LC phases is of course the existence of the helices themselves, and thus the conditions of the duplex denaturation. To probe the relationship between the two transitions, we doped nDNA solutions with EB molecules (*35*), whose fluorescence efficiency is strongly enhanced when they are within the hydrophobic environment of the bases and can thus be used as a marker of denaturation.

We found that the melting temperature T_m, defined as the temperature at which half of the strands are paired, rises with the concentration of nDNA, as is observed for lDNA (*45*). We also find $T_m \sim T_{IC}$: when c is large enough to give the COL phase, then the DNA denaturation occurs along with the LC melting transition to the I phase (the denaturation is spread over 20°, while the I–COL coexistence extends on a few degrees). At lower c we find that $T_m \sim T_{IN} + 10^\circ$, implying that the N phase is more readily disrupted by unpaired strands than the COL phase (*27*).

As already mentioned, a virtually equivalent system of partially paired strands in thermal equilibrium is obtained with mixtures of nDNA where only some of the sequences are complementary, thus able to pair in duplexes, and some are not and thus always remain in the solution as single strands (such a situation can be easily achieved by changing the molar ratio in a mixture of the two complementary sequences CCTCAA AACTCC and GGAGTTTTGAGG). However, in concentrated mixtures of duplex and SS nDNA, the system phase separates into duplex rich LC domains coexisting with a duplex-poor isotropic phase, leading to the physical segregation of complementary chains from non-complementary chains (Figure 4(d)) (*28*). The strong partitioning of the duplex nDNA into the LC domains is confirmed indirectly from PTOM images by measuring the LC fraction, which closely follows the duplex fraction in the system; and directly by fluorescent labelling, attaching a fluorescein (FITC) group externally to the duplexes (*28*). This phase separation, encountered for DS/SS ratio as low as 1/15, is a collective effect of the duplex and SS nDNA, since the duplexes alone would not form liquid crystals in such solutions, their concentration being well below that required for liquid crystal formation.

Two distinct theoretical approaches show that the origin of the nDNA condensation may be the

significant contrast in rigidity between SSs and DSs. Flory, in a mean field theory of the phase behaviour of mixed solutions of rigid rod-shaped polymers and random coil polymers, found N–I phase coexistence of the rigid species preferably into the N and the flexible species preferably into the I phase (*47*). This segregation of species was essentially complete in the high concentration range applicable to our nDNA solutions. Flory's model, wherein the demixing is promoted by the entropic gain associated with the nematic ordering of rigid rods, is appropriate to nDNA solutions if the formation of duplex aggregates precedes the phase separation. If, instead, the formation of living polymers is an effect of the phase separation, an alternative interpretation should be sought. Indeed, a more general mechanism for phase separation of flexible and rigid solutes is provided by depletion attraction, where the segregation of a dense phase of rigid particles enhances the volume available to the flexible solutes and hence the overall entropy (*48*). The compacting of rigid solutes by mean of depletion attractions is well known to promote ordering, as in the case of nematic ordering of rigid rod-like particles mixed with flexible polymers (*49*), the osmotic compression of lDNA into LC phases obtained by adding a hydrophilic flexible polymer to a lDNA solution (*19*), or in the case of protein crystallisation resulting from protein segregation in mixtures with polyethylene glycol solutions (*50*). Interestingly, however, no phase separation was ever detected in mixtures where the FITC group was attached on the 5′ terminal, since FITC with its steric hindrance disturbs end-to-end adhesion. The phase separation is therefore due to a combination of duplex formation and end-to-end stacking, leading to depletion-type forces that arise from the entropy gain of depletants (SSs) upon demixing of solute particles (duplex aggregates, as shown by kinetic studies (*28*)). This manifestation of the depletion interaction is unique in that, above the duplex melting temperature, the depletants and solutes are chemically homogeneous and completely miscible, the difference in molecular size and flexibility enabling a depletion interaction only at temperatures where duplexes can form.

4.1 The role of complementarity and length in phase separation

To gain insight into the potential of the combined stacking, LC ordering and phase separation in the 'purification' of helices, and to test the dependence upon the degree of complementarity of the sequences, we examined a sequence A and an ensemble of sequences B, either fully or only partially complementary to A. By adding controlled mismatches in different positions along the B strands, we could see that LC ordering can tolerate up to four unpaired bases inside the helix, since such mismatches do not affect the overall structure and stability of the helix; conversely, mismatches at the terminals disrupt the N* phase, and the COL phase also disappears when unpaired tails are present. Correspondingly, the phase separation is coupled to the tendency to liquid crystal formation (*51*).

Another approach to address the question on how much the stability of the helices affects the LC condensation is to mix self-complementary sequences of different lengths (6 and 20 bp), to obtain a system of two different helices. In this way, since the melting temperatures of the two strands are appreciably different (around 25°C gap), we also have the possibility of 'switching on' the presence of helices at various temperatures and thus to control the relative amount of the two populations, continuously varying from only SSs, to DSs and SSs, to only DSs. At low T, such a solution forms continuous LC phases in the whole volume, similar to those showed by the single components alone, since the end-to-end stacking mechanism and the LC phases can accommodate them together in the same chains. Instead, within a sample kept isotropic at high T and then cooled just below $T_{\mathrm{I-LC}} \sim 40$°C, the temperature of appearance of the first LC domains, nucleation of a few small domains occurs. At that T, most of the 20 bp helices are already formed, while virtually all of the hexamers are single stranded. Given the high viscosity of the solution (*28*), hundreds of hours are needed for all of the 20 bp helices to diffuse and join in LC domains. Indeed, these domains eventually reach an asymptotic size, around 100 -μm in radius. Figure 6

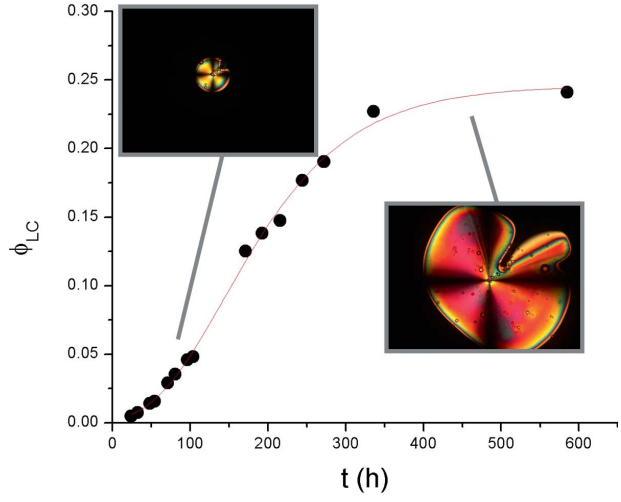

Figure 6. Isolation of longer helices in a mixture of self-complementary hexamers and 20-mers kept at 38°C. After hundreds of hours, the LC fraction ϕ_{LC} reaches a plateau, corresponding to the 20 bp occupancy. The line connecting the data points is a guide to the eye.

shows the time evolution of such domains, together with the fraction of liquid crystals in cell volume, ϕ_{LC}, which agrees within the experimental error with ϕ_{20}, the volume fraction of the 20 bp helices in the sample. We can therefore conclude that the domains are almost entirely formed by 20 bp, and thus the degree of separation is tuned by the amount of paired strands.

5. Prebiotic implications

'*How RNA could possibly have emerged from the clutter without a "guiding hand" would baffle any chemist. It seems possible only by selection, a process that presupposes replication.[. . .] The need seems inescapable for some autocatalytic process such that each lengthening step favours subsequent lengthening. Only in this way could the enormous kinetic obstacle to chain elongation be surmounted. [. . .] Any invoked catalytic mechanism must accommodate the participation of a template, for there can have been no emergence of true RNA molecules without replication*'.

This quotation from a recent book (*52*) by C. de Duve, winner of the 1974 Nobel Prize in Physiology or Medicine and one of the most respected authorities in the research area on the origins of life, clarifies what is probably the main missing link in the chain of the emerging phenomena, leading from the prebiotic 'clutter' to the first living entity (able to replicate and propagate information and, equivalently, to undergo Darwinian evolution, according to NASA definition). According to the so-called RNA-world picture (*53*) (which is the leading model in the field, although disputed), the most primitive form of life was a ribozyme, a ribonucleic enzyme capable of promoting its own polymerisation and thus of self-replicating. However, although the basic constituents of nucleotides may have been available in the prebiotic environment and may even have joined to form oligomers, a critical length of about 50 bp is thought to be necessary to exploit similar catalytic functions (*52*). Clearly, a relatively high concentration is a key requirement for the polymerisation to take place and not to be overcome by hydrolysis and chain degradation. However, high concentration alone is not sufficient to sustain a polymerisation by simple, random ligation of monomers: in a similar manner to the way most of the present enzymes work by geometrical, physical constraints, by keeping close together in active groups and thus enhancing reaction rates, a template mechanism is invoked in prebiotic times to favour ligation of different nucleotides. Various systems, including clays (*54*) and lipids (*55*), have been proposed to act as templates and received promising experimental verification, but conclusive evidence is still to come.

The phenomena described above for nDNA and nRNA, pairing, segregation, stacking and LC ordering of duplexes, could provide at single shot a number of distinctive features: (i) a mechanism of selection, since only paired strands segregate to form LCs; (ii) a locally enhanced concentration of complementary strands and proximity of terminals, up to 10^3-fold; (iii) a liquid crystalline matrix which provides a flexible template, favouring the linear aggregation and possibly the chemical ligation; (iv) an autocatalytic process by which longer and longer complementary sequences are favoured, because they can more easily fit into the aligned environment.

While single nucleotides in solution show a weak propensity to stacking, Guanosine can form quartets (*40*) which are instead strongly prone to the formation of stacked aggregates and LC organisation, being destabilised because of the larger hydrophobic surface. The 'switch' of the stacking advantage, leading to more and more elongated assemblies, is the onset of the reversible and selective H-bonds between the bases.

We can argue that the present nucleic acid molecules emerged as those most capable of self-structuring by a cascade of pairing and stacking into LC structures, enabling chemical ligation, self-replication and hence information storage. To explore the DNA and RNA self-assembly potential and verify our hypotheses, various experiments are currently being pursued on even shorter sequences and mixtures of oligomers with single nucleotides, on random pools of sequences and on non-enzymatic ligation.

6. Conclusion

In conclusion, we have described recent findings on the surprising self-assembly properties of ultrashort fragments of DNA and RNA. Their LC phases share most properties with those of longer DNA strands, but some intriguing features, as demixing of helices from single strands and the much shorter cholesteric pitch, wait for detailed analysis and full understanding.

These results, in addition to confirming DNA as a versatile building block for soft matter research, suggest that the very emergence of nucleotides as information carrying polymers and the string-like structure of nucleic acids could have arisen from the strongly intertwined tendencies of selective pairing, stacking and LC ordering.

Acknowledgements

The work described here would not have been possible without patient and enthusiastic guidance from my supervisor Tommaso Bellini, from Noel Clark and from

the forever missed Michi Nakata. I would also like to acknowledge the collaboration of Marco Buscaglia, Mark Moran, Tommaso Fraccia, Alberto Chierici, Chenhui Zhu, Maggie Hoffman, Taiquitha Robins, Dong Chen, Matteo Salina and discussions with R. Podgornik and P. Collings.

References

(1) Banavar, J.R.; Hoang, T.X.; Maritan, A.; Seno, F.; Trovato, A. *Phys. Rev. E* **2004**, *70*, 041905.
(2) Zandi, R.; Reguera, D.; Bruinsma, R.F.; Gelbart, W.M.; Rudnick J. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15556–15560.
(3) Brown G.H.; Wolken J.J. *Liquid Crystals and Biological Structures*; Academic Press: New York, 1979.
(4) Herzfeld, J. *Acc. Chem. Res.* **1996**, *29*, 31–37.
(5) Hazen, R.M. *Genesis, The Scientific Quest for Life's Origin*; Joseph Henry Press: Washington, DC, 2005.
(6) Lydon, J.E. *Curr. Opin. Colloid Interface Sci.* **1998**, *3*, 458–466; Lydon J.E. *Chromonics*. In: Demus, D., Goodby, J., Gray, G.W., Speiss, H.-W., and Vill, V., Eds; Wiley VCH: Weinheim, 1998, pp. 981–1007; Lydon, J.E. *Curr. Opin. Colloid Interface Sci.* **2004**, *8*, 480–490.
(7) Chiti, F.; Dobson, C.M. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
(8) Israelachvili, J. *Intermolecular and Surface Forces*, 2nd edn; Academic Press: New York, 1992.
(9) Onsager, L. *Ann. NY Acad. Sci.* **1949**, *51*, 627–659.
(10) Adams, M.; Dogic, Z.; Keller, S.L.; Fraden, S. *Nature* **1998**, *393*, 349–352.
(11) Calladine, C.R.; Drew, H.R.; Luisi, B.F.; Travers, A.A. *Understanding DNA—The Molecule and How it Works*, 3rd edn; Elsevier/Academic Press: San Diego, CA, 2004.
(12) Livolant, F. *Physica A* **1991**, *176*, 117–137.
(13) Luzzati, V.; Nicolaieff, V.A. *J. Mol. Biol.* **1959**, *1*, 127–133.
(14) Robinson, C. *Tetrahedron* **1961**, *13*, 219–234.
(15) Rill, R.L.; Strzelecka, T.E.; Davidson, M.W.; van Winkle, D.H. *Physica A* **1991**, *176*, 87–116.
(16) Livolant, F.; Levelut, A.M.; Doucet, J.; Benoit, J.P. *Nature* **1989**, *339*, 724–726.
(17) Livolant, F.; Leforestier, F.A. *Prog. Polym. Sci.* **1996**, *21*, 1115–1164.
(18) Strey, H.H.; Wang, J.; Podgornik, R.; Rupprecht, A.; Yu, L.; Parsegian, V.A.; Sirota, E.B. *Phys. Rev. Lett.* **2000**, *84*, 3105–3108.
(19) Podgornik, R.; Strey, H.H.; Parsegian, V.A. *Curr. Opin. Colloid Interface Sci.* **1998**, *3*, 534–539.
(20) Alam, T.M.; Drobny, G. *J. Chem. Phys.* **1990**, *92*, 6840–6846.
(21) Merchant, K.; Rill, R.L. *Biophys. J.* **1997**, *73*, 3154–3163.
(22) Bolhuis, P.; Frenkel, D. *J. Chem. Phys.* **1997**, *106*, 666–687.
(23) Kramer, E.M.; Herzfeld, J. *Phys. Rev. E* **1998**, *58*, 5934–5947.
(24) Selinger, J.V.; Bruinsma, R.F. *Phys. Rev. A* **1991**, *43*, 2922–2931.
(25) Frenkiel-Krispin, A.; Levin-Zaidman, S.; Shimoni, E.; Wolf, S.G.; Wachtel, E. J.; Arad, T.; Finkel, S.E.; Kolter, R.; Minsky, A. *EMBO J.* **2001**, *20*, 1184–1191.
(26) Levin-Zaidman, S.; Englander, J.; Shimoni, E.; Sharma, A.K.; Minton, K.W.; Minsky, A. *Science* **2003**, *299*, 254–256.
(27) Nakata, M.; Zanchetta, G.; Chapman, B.D.; Jones, C.D.; Cross, J.O.; Pindak, R.; Bellini, T.; Clark, N.A. *Science* **2007**, *318*, 1276–1279.
(28) Zanchetta, G.; Nakata, M.; Buscaglia, M.; Bellini, T.; Clark, N.A. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1111–1117.
(29) Zanchetta, G; Bellini, T.; Nakata, M.; Clark, N.A. *J. Am. Chem. Soc.* **2008**, *130*, 12864–12865.
(30) Van Winkle, D.H.; Davidson, M.W.; Chen, W.; Rill, R.L. *Macromolecules* **1990**, *23*, 4140–4148.
(31) Stanley, C.B.; Hong, H.; Strey, H.H. *Biophys. J.* **2005**, *89*, 2552–2557.
(32) Tombolato, F.; Ferrarini, A. *J. Chem. Phys.* **2005**, *122*, 054908.
(33) Kornyshev, A.A.; Leikin, S.; Malinin, S.V. *Eur. Phys. J. E* **2002**, *7*, 83–93.
(34) Leforestier, A.; Livolant, F. *Biophys. J.* **1993**, *65*, 56–72.
(35) LePecq, J.B.; Paoletti, C. *J. Mol. Biol.* **1967**, *27*, 87–106.
(36) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R.E. *Nature* **1980**, *287*, 755–758.
(37) Redinbo, M.R.; Stewart, L.; Kuhn, P.; Champoux, J.J.; Hol, W.G.J. *Science* **1998**, *279*, 1504–1513.
(38) Davey, C.A.; Sargent, D.F.; Luger, K.; Maeder, A.W.; Richmond, T.J. *J. Mol. Biol.* **2002**, *319*, 1097–1113.
(39) Horowitz, V.R.; Janowitz, L.A.; Modic, A.L.; Heiney, P.A.; Collings, P.J. *Phys. Rev. E* **2005**, *72*, 041710.
(40) Mariani, P.; Mazabard, C.; Garbesi, A.; Spada, G.P. *J. Am. Chem. Soc.* **1989**, *111*, 6369–6373; Davis, J.T. *Angew. Chem. Int. Ed.* **2004**, *43*, 668–698.
(41) Mundy, K.; Sleep, J.C.; Lydon, J.E. *Liq. Cryst.* **1995**, *19*, 107–112.
(42) Lü, X.; Kindt, J.T. *J. Chem. Phys.* **2004**, *120*, 10328–10338.
(43) van der Schoot, P.; Cates, M.E. *Langmuir* **1994**, *10*, 670–679.
(44) Hentschke, R.; Herzfeld, J. *Phys. Rev. A* **1991**, *44*, 1148–1155.
(45) Santalucia, J.; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.
(46) Zanchetta, G.; Nakata, M.; Buscaglia, M.; Clark, N.A.; Bellini, T. *J. Phys.: Cond. Matter* **2008**, *20*, 494214.
(47) Flory, P.J. *Macromolecules* **1978**, *11*, 1138–1141.
(48) Asakura, S.; Oosawa, F. *J. Chem. Phys*. **1954**, *22*, 1255–1256.
(49) Lekkerkerker, H.N.W.; Stroobants, A. *Il Nuovo Cimento* **1994**, *16D*, 949–962.
(50) Annunziata, O.; Asherie, N.; Lomakin, A.; Pande, J.; Ogun, O.; Benedek, G.B. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14165–14170.
(51) Zanchetta, G. Unpublished results.
(52) de Duve, C. *Singularities—Landmarks on the Pathways of Life*; Cambridge University Press: Cambridge, 2005.
(53) Joyce, G.F. *Nature* **2002**, *418*, 214–221.
(54) Ferris, J.P.; Ertem, G. *Science* **1992**, *257*, 1387–1389.
(55) Rajamani, S.; Vlassov, A.; Benner, S.; Coombs, A.; Olasagasti, F.; Deamer, D. *Origins Life Evol. Biosphere* **2008**, *38*, 57–74.