Liquid Crystal Ordering and Isotropic Gelation in Solutions of 4-base-long DNA Oligomers

Tommaso P. Fraccia[1,4], Gregory P. Smith[2], Lucas Bethge[3], Giuliano Zanchetta [1], Giovanni Nava[1], Sven Klussmann[3], Noel A. Clark[2], Tommaso Bellini[1]

1 - Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università di Milano, via Fratelli Cervi 93, I-20090 Segrate (MI), Italy

2 - Department of Physics and Liquid Crystal Materials Research Center, University of Colorado, Boulder, CO, 80309-0390

3 - NOXXON Pharma AG, Max-Dohrn-Str. 8-10, 10589 Berlin, Germany

4 - Dipartimento di Scienze Umane e Promozione della Qualità della Vita, Università San Raffaele di Roma, via di Val Cannuta, 247, I-00166 Roma (RM), Italy

KEYWORDS Liquid Crystals - Self-Assembly - DNA Nanotechnology - Origin of Life - Coaxial Stacking of DNA

ABSTRACT

Liquid crystal ordering is reported in aqueous solutions of the oligomer 5'-ATTAp-3' and of the oligomer 5'-GCCGp-3'. In both systems, we quantitatively interpret ordering as stemming from

the chaining of molecules via a "running-bond" type of pairing, a self-assembly process distinct from the duplex aggregation previously reported for longer oligonuclotides. While concentrated solutions of 5'-ATTAp-3' show only a columnar liquid crystal phase, solutions of 5'-GCCGp-3' display a rich phase diagram, featuring a chiral nematic phase analogous to those observed in solutions of longer oligonucleotides, and two new phases, a columnar crystal, and, at high concentration, an isotropic amorphous gel. The appearance of these phases, which can be interpreted on the basis of features of 5'-GCCGp-3'molecular structure, suggests new assembly motifs specific to ultra-short oligonucleotides.

The motifs of supramolecular ordering of oligonucleotides constitute a field of investigation constantly enriched with new findings, including sequence-directed self-assembly[1], formation of low-concentration hydrogels[2,3], base-stacking driven structures[4,5], various forms of liquid crystalline (LC) long-range ordering and of phase separations[6,7,8]. Here we report that solutions of ultrashort DNA oligomers have a (surprisingly) rich phase diagram at and above room temperature (T) and at concentrations about 1M nucleobases, featuring chiral nematic and columnar LC phases, columnar crystals, and high-density isotropic gel.

Previous investigations described a mechanism for LC ordering of DNA oligomers through a three-step self-assembly process[6]. First, oligomers hybridize into duplexes, which are stable at T below their melting. Second, duplexes form linear aggregates. Such aggregation is mediated by either base stacking of terminal bases in the case of blunt-ended duplexes, or by pairing of terminal sections when duplexes terminate with interacting overhangs. Third, linear aggregates align and form a LC phase. Although these steps mutually strengthen, with LC ordering

enhancing chaining and duplex stability, no LC ordering was previously observed for oligomers too short to form stable duplexes. Indeed, when we observed that LC phases of 6-base-long oligomers were pushed at large DNA concentration (c_{DNA}) and low T [6], we assumed that 4-base-long oligomers would not order, being that their duplexes are unstable at all T above freezing at "ordinary" (< μM) concentrations. Moreover, shorter oligomers imply more flexibility of the aggregates, which is a destabilizing factor for LC ordering[9].

The molecules we investigated and report about in this article are 5'-GCCGp-3' and 5'-ATTAp-3' ("GCCG", "ATTA", preparation and purification details in Material and Methods), partially self-complementary and with interacting overhangs. Their Watson-Crick "running-bond" chaining mode of GCCG is sketched in Fig. 1A, while the one of ATTA is strictly analogous. Quite clearly in this case there is no distinction between duplex formation and linear aggregation. Rather, hybridization and chaining necessarily take place in a single continuous process, in which the bonding of each GCCG or ATTA molecule to an existing chain (Fig. 1B) and the merging of different chains (Fig. 1C) are characterized by a free energy gain. This condition leads to equilibrium states characterized by a distribution of lengths, ranging from single dissociated oligomers to multi-molecular aggregates. As typical for "living polymers"[10], such distribution is expected to be exponential, with characteristic length depending on interaction strength and T, as discussed below.

RESULTS

Pairing modes of ATTA and GCCG and their thermal stability

In Figs. 1A-C we sketch the various interactions that contribute to the formation of the linear aggregates. Indeed, besides the contributions to the free energy from the formation of

quadruplets, i.e. pairs of adjacent Watson-Crick bonds between distinct strands whose free energy is typically determined in the frame of nearest-neighbor (NN) model[11] (double black and red arrows in Fig. 1A), important additional contributions come through "coaxial stacking", i.e. the stacking interactions of chemically disconnected but physically contacting paired bases (green arrows in Fig. 1A), and through "dangling ends", i.e. the stacking of an unpaired single base adjacent to a C-G or A-T pair (in the position marked by a pink arrow in Fig. 1A). Coaxial stacking and dangling end are here essential and make the binding of the chains much stronger than that in the formation of lone pairs of GCCG and ATTA. It should also be noticed that, because of its low molecular weight, the molar concentration of GCCG and ATTA where LC phases are found is in the range 0.2 M – 0.6 M. At these large concentrations molecular associations with a binding coefficient as low as a few k_BT such as those here considered, become relevant.

A straightforward approach to estimate the strength of ATTA and GCCG base-paired chains is to measure their thermal melting through the fluorescent emission $I_F(T)$ of ethidium bromide (EtBr), a fluorophore whose fluorescence quantum yield markedly increases upon its intercalation between paired bases of DNA. We find that in both systems, EtBr fluorescence decay reveals a single, broad thermal transition occurring at a T depending on c_{DNA}, which at the highest concentrations reaches 60°C, a remarkably large value. While normally EtBr melting curves are interpreted in terms of duplex melting, here this notion should be revisited because of the ATTA and GCCG chaining mode that leads to a distribution of aggregate lengths. In this case, instead, the fraction of bonded bases revealed by I_F provides information on the mean chain length. To this aim, we adopted the description of linear aggregation provided by recent models that enable calculating the chain length distribution cylinders having adhesive bases with

interaction strength ΔG [13,14]. We thus model our chains as cylinders whose size is that of 2-base-long duplexes, since this corresponds to the lengthening of the aggregates for each addition of a 4mer. Within this assumption we analized our $I_F(T)$ data and determined the free energy gain ΔG that favors the chaining of ATTA (ΔG_{ATTA}) and GCCG (ΔG_{GCCG}). Since these values are determined from conditions in which the monomeric ATTA and GCCG are a small fraction of the population (making the process in Fig. 1B of minor importance), we argue that the value of $\Delta G_{ATTA}(T)$ and $\Delta G_{GCCG}(T)$ that we extract represent the inter-chain bonding free energy sketched in Fig. 1C. $I_F(T)$ data and analysis are reported and extensively discussed in Supplementary Information, S.I.

The values obtained for $\Delta G_{ATTA}(T)$ and $\Delta G_{GCCG}(T)$ (in table S1) can be compared with those calculated with the well-established NN thermodynamic parameters for DNA stability to obtain new information on coaxial stacking. Indeed, while the thermodynamic stability of quadruplets and the contribution of dangling ends is rather well determined, the value of coaxial stacking has been studied less and with quite different outcomes[12,13]. Therefore, by subtracting from $\Delta G_{ATTA}(T)$ and $\Delta G_{GCCG}(T)$ the effects of quadruplets and dangling ends, we obtained a new determination of the thermodynamic parameters of coaxial stacking. Namely, for nicks A-A/TT we find ΔH^{AA}_{coax}= -3.7 kcal/mol and ΔS^{AA}_{coax}= -4.8 cal/(mol K), while for nicks G-G/CC we find ΔH^{GG}_{coax}= -12.6 kcal/mol and ΔS^{GG}_{coax}= -32.7 cal/(mol K). These values are included in the wide interval between the results of two studies reported in literature[12,13] (see S.I.).

Phase behavior of ATTA and GCCG

We have studied the phase behavior of solutions of GCCG and ATTA without added salt at concentrations up to $c_{DNA} \approx 800$ mg/ml, above which samples are difficult to handle and

homogeneity hard to obtain. Exploration by polarized transmission optical microscopy (PTOM) of the phase behavior of solutions in thin flat cells enabled identification of domains in the c_{DNA}-T plane exhibiting birefringent textures typical of liquid crystal ordering as well as domains in which we found coexistence of birefringent and isotropic phases. The PTOM textures observed in ATTA cells indicate that this system can order into a columnar (COL) LC phase (Fig. 1D) in the c_{DNA}-T range reported in Fig. 1F. The PTOM textures of GCCG indicate a richer phase behavior, featuring chiral nematic (N*) LC ordering and the formation of a more rigid ordering that we identify as a columnar crystal phase (COLX), Fig. 1E. The domains for the N* and COLX phases are reported in Fig. 1G. In both ATTA and GCCG phase diagrams black, blue and red symbols indicating isotropic I, N* and COL/COLX phase, respectively. Open symbols indicate N*-ISO and COL/COLX-ISO coexistence. Green dots mark N*-COLX coexistence. In both phase diagram, LC ordering, although in coexistence with the isotropic phase, is remarkably stable, found at higher temperatures and lower concentrations than in longer blunt-ended oligomers (6-10mers). The phase behavior of GCCG shows a few remarkable features not found previously in other DNA oligomers nor in the ATTA phase diagram: (i) the appearance of an isotropic phase which replaces LC ordering at high c_{DNA}; (ii) a resulting bell-shaped structure of the LC region in the phase diagram; (iii) a crystalline columnar phase, COLX, replacing the COL LC phase found in other DNA oligomers and in ATTA solutions. These topics are analyzed and discussed in the following sections.

Quantitative analysis of the phase diagrams

Having determined the association free energies ΔG_{ATTA} and ΔG_{GCCG} involved in the molecular assembly puts us in position of quantitatively compare the phase diagrams in Figs. 1F and 1G

with the prediction of theoretical models for the collective behavior of cylindrical monomers undergoing base-to-base aggregation[14,15].

First, there is a qualitative difference between the phase diagram of ATTA, in which the N* phase is absent, and the one of GCCG, where instead it is found. According to ref.[14] the nematic phase requires stronger bonds than the COL phase. This is shown in Fig. 1H, where we report the phase diagram computed in ref.[14] for cylinders with axial ratio 0.5, the closest to our condition among the values considered in that paper. As visible, the range in which the nematic phase is stable is small and bonded in energy. This is a consequence of the fact that the N* ordering can be found only at concentrations lower than those of the COL phase, where at higher T, the mean aggregation length is not enough to induce LC ordering. Specifically, Fig. 1H indicates that the N* phase requires end-to-end interactions to be larger than a value ΔG_{MIN} that, for an axial ratio of 0.5, is predicted to be $\Delta G_{MIN}/k_BT \approx 9$. Correspondingly, the concentration of maximum stability for the nematic phase is expected to be at cylinder volume fraction of about 0.37, roughly equivalent to c_{DNA} = 380 mg/ml (see S.I.). Remarkably our observed phase diagrams nicely agree with this prediction. At T = 30 °C, the largest T at which the N* phase is found, $\Delta G_{GCCG} \approx 7.3\ k_BT$, which can be assumed to be the ΔG_{MIN} for our experimental system (blue dashed line in Fig. 1H), and $c_{DNA} \approx 370$ mg/ml. Instead, even at T = 0 °C, the lowest T here explored, $\Delta G_{ATTA} \approx 7\ k_BT$, insufficient for the appearance of nematic ordering. The intervals of energies and volume fraction of equivalent cylinders explored in the experimental phase diagram are $5\ k_BT < \Delta G_{GCCG} < 12\ k_BT$ and $\Phi > 0.23$ for GCCG, and $4.4\ k_BT < \Delta G_{ATTA} < 7\ k_BT$ and $\Phi > 0.3$ for ATTA, respectively highlighted by blue and yellow contours in Fig. 1H.

Furthermore, the same model of Ref.[15] that we used to extract the free energies, enables calculating the region of the phase diagram where LC ordering is expected. Specifically, the

model enables determining the ISO side of the ISO-N phase coexistence (which can be extended to ISO-LC since the model does not include a distinction between N and COL phases). We thus performed this calculation using ΔG_{GCCG} and ΔG_{ATTA}. The results are plotted in Fig. 1F and Fig. 1G as purple lines in the concentration interval of confidence for the model (200-400 mg/ml, further discussion in the SI). The agreement is excellent in both cases.

The phase diagram of GCCG in Fig. 1G differs from the one in Fig. 1H by the presence, at large concentrations, of an isotropic phase which appears to be a specific feature of this system, since upon dehydration DNA LC phases quite generally transform into more ordered, typically crystalline, states. In the case of GCCG, however, upon increasing c_{DNA} the COLX phase transforms into an isotropic state, here indicated as ISO2. This was repeatedly tested in evaporative cells at constant T, in which as c_{DNA} grows, the system horizontally crosses the phase diagram to eventually enter the ISO2 state which is maintained up to complete dehydration (see S.I.). This was also tested by preparing a sample in a 1.5 mm diameter glass capillary by inserting a 5.6 mg of lyophilized GCCG at its bottom and serially adding water to span c_{DNA} through the whole range of interest. At each concentration the samples were equilibrated, observed at different T and gently centrifuged at given Ts. Pictures of the capillaries before and after centrifugation are reported in Fig. 1H for the values of c_{DNA} marked in Fig. 1G by vertical lines and labels 1-4. Once separated by centrifugation, the coexisting phases remained indefinitely stable in the tubes at the sedimentation temperature, indicating equilibrium state. We confirmed that in coexistence densities are ordered as $c_{DNA}(ISO) < c_{DNA}(N^*) < c_{DNA}(COLX) < c_{DNA}(ISO2)$. By centrifuging capillary 1 (350 mg/ml) at 20 °C, the N* phase sediments at the bottom, while the ISO phase floats on top. A similar result is found in capillary 2 (c_{DNA} = 500 mg/ml) at 35 °C for the COLX phase. The behavior is reversed in capillary 3 (c_{DNA} = 600 mg/ml)

at 35 °C in which centrifuging leads to the sinking of the ISO2 phase and floating of COLX. Capillary 4 remains featureless under centrifugation, compatible with a homogeneous ISO2 phase.

The presence of isotropic fluid on both sides of the diagram makes this phase diagram analogous to those reported for surfactant solutions where often lamellar and hexagonal phases are limited on both larger and smaller concentration by isotropic micellar solutions[16]. This topology of the phase diagram in which the ordered phase is surrounded by isotropic fluid is also found for binary mixtures in which mixing is more favored in the ordered phase than in the disordered one such as what found in mixtures of enantiomers in which the liquidus curve has a maximum[17]. The point of maximum stability, i.e. the vertex of the bell-shaped envelop of the ordered phase (COLX in our case), is a first order transition in which the transformation is discontinuous in T (with no phase coexistence), as in a single component system.

In our case, however, the diagram shows some degree of non-ideality since the COLX and the COLX-ISO coexistence do not merge at a single point: in the T interval 45-50 °C the COLX phase can be found only in coexistence with the ISO phase, while 45 °C is the largest T at which an homogeneous COLX phase can be found. This is clearly incompatible with a two-component system, but it is easily understood as an effect of minor residual impurity, which are revealed by osmolarity measurements as discussed in the S.I. This non-ideality is also revealed by measuring the concentration of coexisting ISO and LC phases. This measurement was performed by preparing three narrow capillaries, which were let equilibrate and then centrifuged: a c_{DNA} = 350 mg/ml GCCG capillary centrifuged at T = 22°C, a c_{DNA} = 600 mg/ml GCCG at T = 32°C, and a c_{DNA} = 600 mg/ml ATTA at T = 22°C. These capillaries were cut at the meniscus and the concentrations measured by UV absorption. Results are shown in Fig. 1F and Fig. 1G as larger

dots over a grey shading. In all cases the difference in c_{DNA} between the two phases is of the order of 50 mg/ml, in line with our previous observations[7,8] and theory[14]. This value is markedly smaller than the width of the coexistence regions in Figs. 1F and 1G, again revealing the presence of an additional component which partitions between the two phases, effectively widening their coexistence.

Characteristics of the phases of GCCG

Chiral nematic (N) phase*

The GCCG N* phase is characterized by a left-handed helical precession of the director and by helical pitches in the range 1-5 μm and growing with T and with c_{DNA}. Handedness and value of the N* pitch are known to be the outcome of a subtle balance of competing forces[18]. The handedness as well as the T and c_{DNA} dependence of the GCCG N* phase are the same as those observed in the N* phase of DNA duplexes when the nematic ordering develops at low enough DNA concentration, typically when c_{DNA}(ISO-N*) < 650 mg/ml[19] as in this specific case. In this condition the side-to-side distance between parallel helices is large enough so that the N* chirality is dominated by the electrostatic interactions between the phosphate chains favoring left-handed phase chirality [18–20]. Further data and discussion in the S.I.

Columnar crystal (COLX) phase

Fig. 2A shows polarized transmission light microscope pictures of the COLX phase in coexistence with the ISO phase. The hallmark of the columnar phase in DNA solutions, first identified by Livolant and co-workers, are the conformal domains in which the local columnar axes lie on set of nested circles, a structure indicating repulsive interactions between adjacent columns that allow for their free sliding along the column axis[21]. This is the columnar texture

also typically found in LCs of oligomeric DNA and RNA as well as in ATTA (Fig. 1D). By contrast the columnar texture found in GCCG indicates that the optical axis is uniform in each domain, as confirmed by the simultaneous extinction of the transmitted light from the whole domain (Fig. 1B). Indeed, despite the many observations of this phase in GCCG samples, we have never observed the typical fan-shaped textures. This indicates that bend-type distortions of the director, allowed by the symmetry of the columnar phase, the flexibility of the columns, and their ability to slide past one another, are in GCCG instead suppressed.

We measured the birefringence Δn of the COLX GCCG domains and found $\Delta n/c_{DNA} \approx 6.2 \pm 0.4\ 10^{-5}\ l/g$, consistent with the birefringence of the columnar phase measured in solutions of DNA oligomers[19] and polymers[22,23]. This indicates that paired bases in the COLX phase are ordered with their normal along the optical axis as in the conventional DNA columnar phase. Since any other arrangement would decrease the degree of orientational order of the bases and thus significantly lower the birefringence, we understand the COLX phase as a columnar phase with additional side correlation among the columns forbidding bend deformation. In other words, we understand the COLX phase as a crystal-columnar phase.

Given the symmetry of the system, side-side interactions between columnar aggregates of GCCG and ATTA must reflect its 2-base-long structural periodicity, as sketched in Fig. 2C. Such periodicity might cause a 2-base-long modulation of shape, and/or flexibility, and/or charge distribution. A possible source of these are transient unbindings of terminal pairs at each coaxial stacking position. Quite clearly any of these modulations might induce longitudinal locking between neighboring columns, their loss of sliding freedom driving the COL phase to crystallize. A simple demonstration that column periodicity promotes stability of the GCCG COLX phase is offered by observing the effect of adding EtBr to the DNA solutions ([EtBr] ≈ [GCCG]/2). Since

EtBr randomly intercalates between paired bases, it acts as a local stretcher of the helical structure, thus disrupting its periodicity, as sketched in Fig. 2D. Observations of LC textures by polarized light microscopy systematically confirm that the insertion of EtBr in the system eliminates the crystal character of COLX, transforming it into a standard bend-prone columnar LC phase, as shown in Fig. 2E - F. The optical axis of the GCCG - EtBr phase (green lines in the figure), corresponding to the direction of low refractive index (i.e. along the duplex axis), is free to bend, enabling the typical columnar fan-shaped LC textures.

The formation of COLX in GCCG can be understood as the result of two effects. First, the onset of slide-forbidding correlations is made simpler in ultrashort oligomers. A simple estimate of the entropic penalty involved in the longitudinal locking of the columns is given by $\Delta S_L \approx k_B \ln(\delta/\Lambda)$, where $\Lambda \approx 6.6$ Å is the period along the GCCG aggregate and δ is the sliding freedom of locked columns, likely of the range of one inter-base distance $\delta \approx 3$ Å. This estimate suggests that the locking of GCCG occurs for inter-aggregate longitudinal interaction of the order of $\Delta G_L \approx 1\ k_BT$. In the case of longer LC-forming oligomers Λ becomes larger, and with it the entropic penalty to longitudinally lock the columns. Second, we find the locking mechanism is to be present in GCCG but not in ATTA. This could be a consequence of larger shape modulations of GCCG. A further possibility to account for this difference is the richer arrays of modes of interaction available for guanosines (G), which are known to be able to form H-bonds between molecular positions not involved in conventional Watson-Crick pairing, such as those between Gs in the so-called G-quadruplexes[4] (sketched n Fig. 2G). It is thus possible that as G terminals transiently unbind from the duplex aggregate, they bond with Gs which are part of neighboring columnar aggregates, contributing to the longitudinal locking of the columns.

Dense isotropic phase

One of the most intriguing features of the GCCG phase diagram is the LC phases being delimited by ISO phase on both the low density and the high-density sides. While we certainly expect the presence of an ISO phase at low c_{DNA} and high T where the degree of GCCG association is lower, the presence of a large c_{DNA}, low T ISO2 phase is specific for GCCG, since all other oligos we have considered in our previous studies develop, in that limit, some form of birfringent crystalline structure.

The ISO2 phase appears to change continuously and directly into the ISO phase while remaining optically isotropic, which makes this transition difficult to observe optically in cells and capillaries at large c_{DNA}. Indeed, we never could detect coexistence between distinct isotropic phases, for example by preparing a flat cell at c_{DNA} = 700 mg/ml and carefully examining it upon cycling T (see S.I. and Supplementary Movies).

To gain a better sense of the effect of cooling on in samples at large c_{DNA}, we performed Dynamic Light Scattering (DLS) on the sample contained in capillary 4 of Fig. 1H after having dissolved in it a small amount of colloidal particles having 100 nm diameter. Due to the small size of the capillary, we couldn't perform a reliable study as a function of the scattering vector. However, we could measure the intensity correlation functions $g_2(\tau) = \langle I_s(t) I_s(t+\tau)\rangle_t / \langle I_s(t)\rangle_t^2$ at fixed geometry (scattering angle ≈ 90°) at various T. Fig. 2H shows the squared field correlation function $|g_1(\tau)|^2 = (g_2(\tau)-g_2(\infty))/(g_2(0)-g_2(\infty))$ for three different T. At large T (T = 65 °C), where at visual inspection the solution appears fluid (see S.I.), we observe a single polydisperse decay indicating some degree of colloidal aggregation. The characteristic time of the fastest components of the decay suggest that the viscosity of the solution, even at this large T where we don't expect any structure, is up 10 times larger than pure water. This is expected, given the large c_{DNA}. As T decreases at and below values at which the solution appears to be a gel

(see S.I.), a second feature appears in the correlation at shorter times, whose amplitude grows upon while its characteristic time slows down. This feature must be a manifestation of growing GCCG concentration fluctuation related to the increased intermolecular interactions, although no clean description of this is available at this stage. In the same range the main, slower feature markedly slows down (arrows in the figure), reaching the upper limit of measurable correlation times. In the interval 65°C – 35 °C, the relaxation due to colloidal diffusion slows down more than three orders of magnitude, indicating an analogous growth of the viscosity. For even lower T, the scattered light shows the typical features of non-ergodic systems, such as a low contrast of the time-averaged correlation functions.

This simple observation of the dynamics of scattered light indicates that the transition to ISO2 is associated to a strong increase of viscosity and kinetic attest, a behavior that closely resembles the continuous gelation observed in DNA hydrogels[24]. While determining molecular structure of ISO2 will require further structural analysis such as x-ray diffractometry, it is worth noticing that it is indeed expected that the liquid crystal ordering of a molecule such as GCCG would become unstable at large enough concentration. Because of the large fraction of G in the sequence, as c_{DNA} increases structure based on G-quadruplexes (Fig. 2G) become more stable than those based on Watson-Crick pairing. This is because the equilibrium concentration of G-quadruplexes depends on $[G]^4$, the fourth power of the concentration of Gs [25], while Watson-Crick pairs depend on c_{DNA}^2. Moreover, G-quadruplets can stack on each other and the equilibrium concentration of two stacked G-quadruplexes depends on $[G]^8$. Consider, for example, an aggregation motif involving G-octuplets as the one sketched in Fig. 2L, consisting of pairs of stacked G-quadruplexes formed by the terminal Gs of eight distinct GCCG molecules. This structure is expected to become energetically favorable over Watson-Crick running-bond linear aggregates at

room T for concentrations larger [G] ≈ 1 M (c_{DNA} ≈ 650 mg/ml), that is where we indeed found the COLX-ISO2 transition. In this evaluation we adopted the thermodynamic parameters for G-quadruplexes given in Ref. [25] at room T (further discussion in S.I.). Moreover, a simple estimate of the volume occupied by the repeating unit in Fig. 2L is of the order of 10 nm^3, matching $8\ell/N_A$ ≈ 12 nm^3 (where ℓ indicates liters), the volume available to each unit calculated on the basis of the GCCG concentration at [G] ≈ 1 M.

CONCLUSIONS

In this study we investigated and quantitatively analyzed the phase behavior of the DNA oligomers ATTA and GCCG, whose sequence favor a running-bond staggered linear aggregation. Despite their expected weak bonding, these 4-base-long oligomers are here reported to feature a rich phase diagram including N* and COL liquid crystal phases analogous to the one previously reported for longer DNA oligomers, and COLX and ISO2 molecular ordering, not observed previously in DNA solutions.

By adopting a recently developed model for the end-to-end assembly and liquid crystal ordering of cylindrical particles, we find a very good agreement between the thermal stability of the linear aggregates measured at various DNA concentration and the phase boundary marking the appearance of liquid crystal ordering. Our analysis also leads to a new determination of the strength of the contribution of coaxial stacking to the duplex stability. Indeed, because of their heavily segmented structure, the aggregation of GCCG and ATTA appear ideal systems to provide a more solid deduction of the thermodynamic parameters for coaxial stacking, so far known only rather loosely.

The columnar crystal COLX and the dense isotropic ISO2 phases are found in GCCG solutions and are here interpreted as a possible effect of the interactions between guanosines. Indeed, the H-bonding between Gs could indeed provide additional interactions between distinct GCCG double-helical aggregates, thus suppressing their freedom to slide along the columnar axis. Moreover, the thermodynamic stability of G-quartets and octets determined in previous studies supports the notion that at large enough concentration, the system should prefer aggregation states alternative to the linear aggregates, as experimentally found.

Overall, the discovery of LC phases in DNA 4mers complements previous observations of LC phases in solutions of random-sequence oligomers[8] and in multicomponent mixtures[7], confirming the remarkable capacity of nucleic acids to develop long-range fluid ordering. We previously speculated that these modes of self-assembly could have played a role in the prebiotic appearance of nucleic acids[26]. The results described here strengthen this notion and open to the possibility of the existence of ordered phases in solutions of even shorter oligomers.

METHODS

Synthesis and purification

DNA 4-mers were assembled on an Äkta10 Oligopilot synthesizer (Amersham Biosciences; GE Healthcare, Freiburg, D) in a 1.2 mL fixed volume column using standard DNA phosphoramidites. Syntheses were started on dT base loaded CPG, pore size 600 Å, loaded to 70μmol/g (Prime Synthesis, Aston, PA, USA). The introduction of the 3'-phosphate modification was achieved by coupling Chemical Phosphorylation Reagent (CPR, Link Technologies, Bellshill, UK). An capping-oxidation-capping cycle was applied. Cleavage and deprotection was achieved by 90min treatment with 40% aq. $MeNH_2$ solution. This treatment not

only cleaves the oligonucleotide from the solid support and removes all protective groups, but also releases the 3'dT and thus yielding 3'-phosphate modified 4-mer[27]. After cleavage and deprotection the product was harvested by filtration and 50% EtOH wash. The filtrate was evaporated to dryness and then reconstituted in 1 mL water. To this solution, 100 μL of a 3 M NaAc solution was added and the oligonucleotide was precipitated by addition of 14 mL MeOH/BuOH (1/3, v/v) to remove acrylonitrile and acylamides from cleavage and deprotection. The precipitate was collected by centrifugation and decantation and washed with acetone. This precipitation procedure was repeated. The final pellet was dissolved in 5 mL water and further purified by size exclusion chromatography (SEC, Sephadex G25, GE Healthcare Europe, Freiburg, D) to remove excess salt and 5'phospate-dT. This process yielded 796 OD (26.3 mg, 20.9 μmol) 5-GCCGp-3' and 888 OD (29.3 mg, 23.4 μmol) 5'-ATTAp-3'. Purity and identity was confirmed by IEX- and RP-HPLC and LC-MS (ESI-).

Sample preparation

Stock solution of ATTA and GCCG were prepared in MilliQ water at c_{DNA} = 100 mg/ml (~85 mM) and stored at 4°C. Three classes of samples were studied.

1) Thin flat cells were prepared by progressively depositing 0.5 - 1 µl droplets of the stock solutions on flat glass (3 x 1.25 x 0.09 cm) and letting them dry until the required concentration is achieved. The cells were then closed, sealed with a fluorinate oil and glued. Celle gap was provided by 20 µm silica spacers. Cell thickness and c_{DNA} were measured by performing microscope-based interferometry in different spots of the cell, i.e. across the fluorinated oil, and across DNA solutions in isotropic phase (at large T).

2) A Vitrocom round capillary of 1.5 mm inner diameter was prepared by inserting 5.6 mg of lyophilized GCCGp in the bottom of the tube and serially adding water, spanning a total volume in the range 6-20 µl.

3) Thinner capillaries, 0.6 mm inner diameter, were prepared by sucking 2 - 4 µl of DNA solution at the desired concentration and flame sealed. After DNA sedimentation upon centrifugation at $3.5 \cdot 10^3$ g in a T controlled Eppendorf 5702RH centrifuge, the capillaries were cut at the meniscus between phases. The content of the resulting capillaries sections was diluted in 50 µl water and measured by 260 nm absorption with Evolution 300 UV-Vis spectrophotometer (Thermo Scientific). Separation of coexisting phases from centrifuged thin capillaries were used to measure in each phase DNA concentration – via UV absorbance – and the osmolarity – via freeze point depression.

Optical microscopy, osmolarity measurements, light scattering measurements

Phases observations and fluorescence measurements were obtained with polarized optical microscopes TE300 and Abbe Optiphot 2 (Nikon) equipped with Nikon DS-U1 color camera and Jenoptik b/w fluorescence camera. Temperature was controlled by Instec TSA02i hot and cold stages.

Dynamic light-scattering measurements were performed on a ST100 SciTech Instruments apparatus, customized to measure samples with small volume. Measurements were performed with a 532-nm solid-state laser source at 60° scattering angle.

Osmolarity was measured with a Gonotec Osmomat 030 (Gonotec, Berlin, D) freezing point osmometer. To perform the measurements, samples were adjusted to 5mM concentration in water and a 50μL aliquot was used for the measurement.

FIGURES

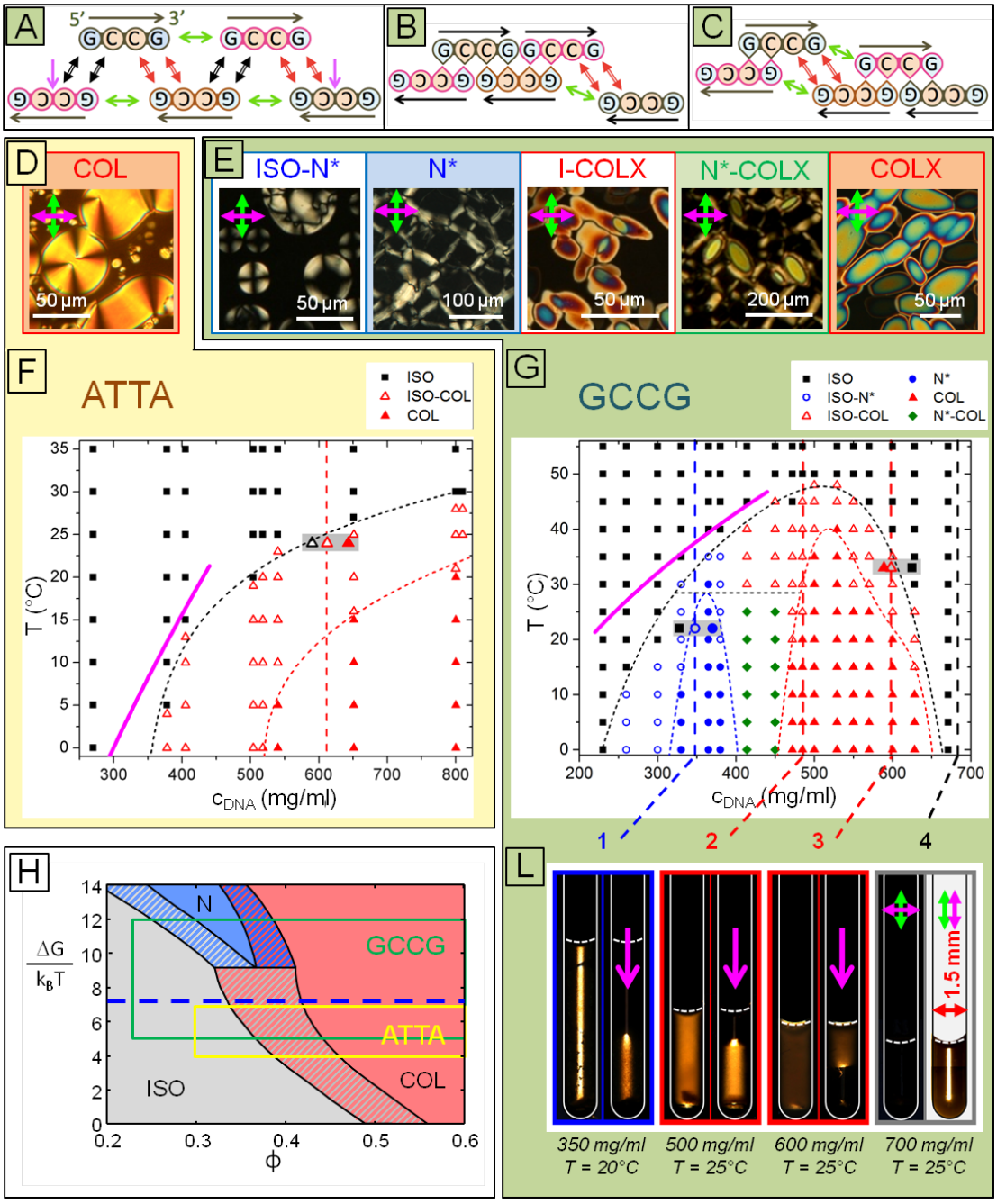

Figure 1. **Modes of assembly and phase diagram of the 4mers 5'-GCCGp-3' and of 5'-ATTAp-3'.** A: Pairing and stacking interactions at play in the stability of the GCCG aggregates: Watson-Crick GC/CG quartets (black arrows) and CG/GC quartets (red arrows), coaxial stacking (green arrows), dangling end

stacking (between the bases as marked by pink arrows). B-C: Interactions at play between an aggregate and a monomer (B) and between two aggregates (C). D-E: Pictures by polarized transmission optical microscopy of thin cells of ATTA (D) and GCCG (E) at T = 10°C and distinct concentrations (from left to right c_{DNA} = 650 mg/ml, c_{DNA} = 260 mg/ml, c_{DNA} = 365 mg/ml, c_{DNA} = 450 mg/ml, c_{DNA} = 500 mg/ml and T = 25°C, c_{DNA} = 570 mg/ml). F-G: c_{DNA} – T phase diagrams for ATTA (F) and for GCCG (G). Symbols indicate: ISO and ISO2 phase (solid black squares), ISO-N* coexistence (empty blue circles), N* phase (solid blue circles), ISO-COL and ISO-COLX coexistence (empty red triangles), N*-COLX coexistence (solid green diamond), COL and COLX phase (solid red triangles). Thin dashed lines guide the eyes to identify the phase boundaries (black - ISO, blue - N* and red - COL/COLX). Magenta lines mark the ISO side of the ISO-LC phase coexistence according to the model of Ref. [15] implemented with ΔG_{ATTA} and ΔG_{GCCG} obtained from melting temperature measurements. Vertical lines and 1-4 labels mark the c_{DNA} where capillaries were prepared and investigated. Symbols over a grey shading indicate the c_{DNA} of the two coexisting phases measured after cutting centrifuged capillaries at the meniscus. H: Φ - $\Delta G/k_BT$ phase diagram for DNA duplexes with L/D = 0.5 from ref.[14], solid contours highlight the volume fraction - energy regions investigated in the experimental phase diagrams of ATTA (yellow) and GCCG (green), blue dashed line indicate the experimental $\Delta G_{MIN}/k_BT \approx 7.3$ for the nematic ordering. L: Pictures through crossed polarizers of the larger capillary before and after centrifugation (magenta arrow) showing ISO-N* (1), ISO-COLX (2), COLX-ISO2 (3) and ISO2 (4) at c_{DNA} and T as indicated.

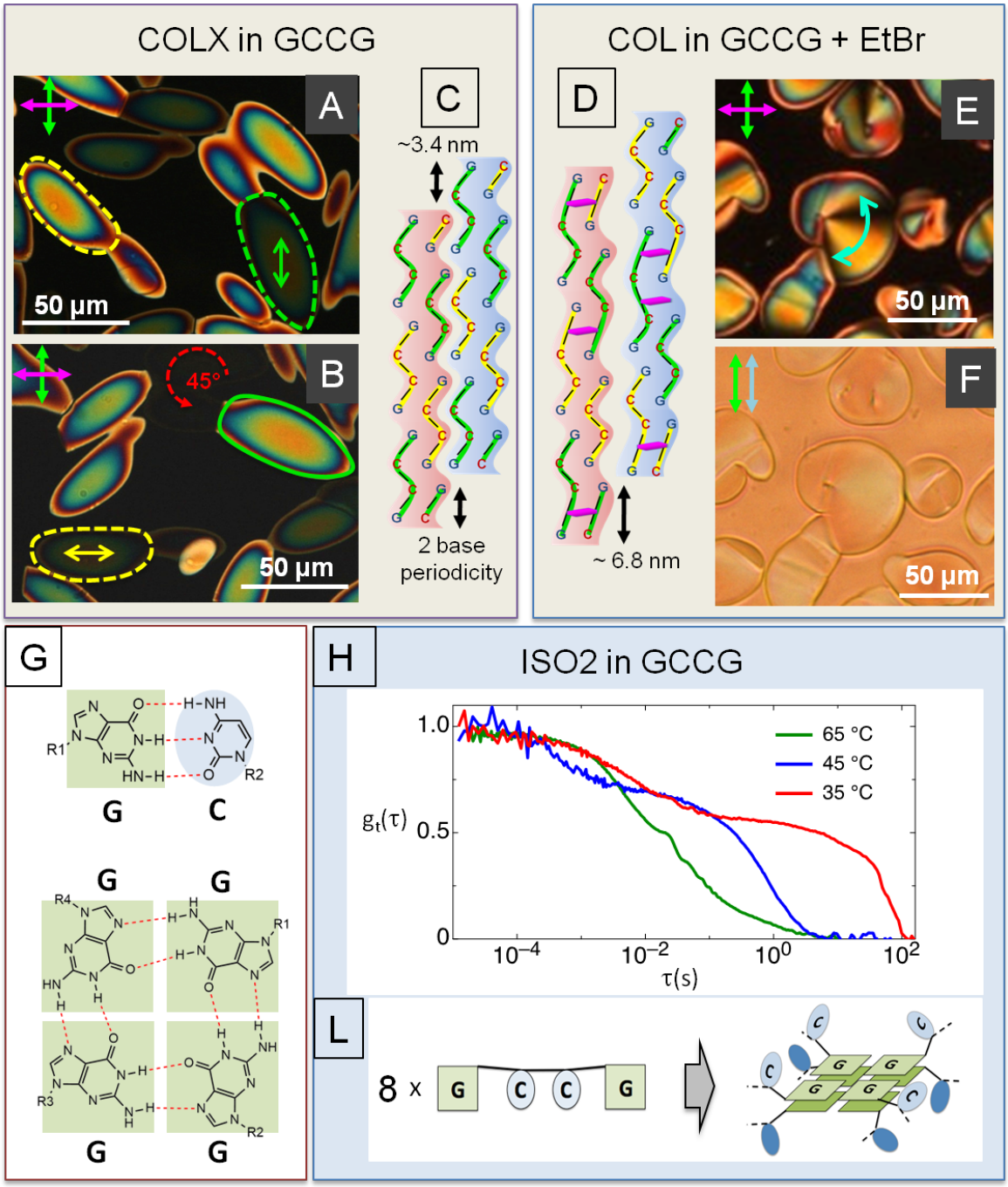

Figure 2. **Description of the liquid crystalline phases of GCCG.** A-B: Polarized optical microscopy pictures in crossed polarizers of the same portion of a thin cell with COLX domains. Domains marked with a yellow and green dashed line contour show extinction and maximum brightness upon rotating the cell of 45°. Yellow and green arrows indicate the optical axis (axis of least refractive index), uniform within each domain. C: Sketch of two GCCG aggregates showing the 2-base long periodicity which

enables longitudinal locking. The shape modulation in the drawing is intended as indicating a steric modulation or a modulation of flexibility or of interaction strength between the bases of distinct aggregates. D: The periodicity and locking mechanism is disrupted by the random intercalation of EtBr (pink segments). E: Polarized optical microscopy pictures of a domain of GCCG+EtBr showing fan-shaped textures typical of COL LC ordering (light blue arrow indicates the distorted optical axis). F: Transmission microscopy picture between parallel polarizers showing that the absorbance from EtBr is modulated following the direction of the optical axis, confirming intercalation and thus orientational order. G: Guanine (green shading) and Cytosine (blue shading) form Watson-Crick bond (left-hand side) and quadruplex bond (right-hand side). H: Time-averaged intensity correlations of the scattered light (λ = 532, angle ≈ 90°) from a thin capillary in ISO2 phase (c_{DNA} = 700 mg/ml) at different T (see legend). The plot shows the quantity $g_t(\tau) = \langle I_S(t)I_S(t+\tau)\rangle_t/\langle I_S(t)\rangle_t^2$, I_S being the scattered intensity. L: Sketch the structural element possibly at the basis of ISO2 ordering: terminal Gs of 8 distinct GCCG molecules assemble in two stacked G-quadruplexes. At large enough GCCG concentration, this structure is expected to be more stable than the Watson-Crick duplexing.

ASSOCIATED CONTENT

Supporting Information

Additional information on the aggregation model, aggregates thermal stability, coaxial-stacking energy values, prediction of the isotropic-LC transition, osmolarity measurements, G-quartets vs. duplex equilibrium and further data on phase characterization are presented in the Supporting Information, which includes Figures S1-S8 and Table 1. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* BIOMETRA Dpt., Università di Milano, via Fratelli Cervi 93, 20090 Segrate (MI). Email: tommaso.bellini@unimi.it

Author Contributions

T.B., N.A.C., G.Z., T.P.F. and G.P.S. conceived the experiment; T.P.F. measured phase diagrams and thermal stability; T.B. adapted the aggregation model to analyze the data; G.N. performed the DLS measurements; L.B. and S.K. synthesized the oligomers and performed osmolarity measurements; T.B., N.A.C., G.Z. and T.P.F. wrote the manuscript. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported by the PRIN Grant Program of the Italian MIUR Ministry and by NSF MRSEC Grants 0820579 and 1420736, and NSF Biomolecular Materials Grant 1207606.

ACKNOWLEDGMENT

We acknowledge Carsten Tschierske , David M. Walba and Marco Todisco for useful suggestions and stimulating discussions. This work was supported by the PRIN Grant Program of the Italian MIUR Ministry and by NSF MRSEC Grants 0820579 and 1420736, and NSF Biomolecular Materials Grant 1207606.

ABBREVIATIONS

GCCG, 5'-GCCGp-3'; ATTA, 5'-ATTAp-3'; G, guanosine; EtBr, ethidium bromide.

REFERENCES

(1) Jones, M. R.; Seeman, N. C.; Mirkin, C. A. Programmable Materials and the Nature of the DNA Bond. *Science (80-.).* **2015**, *347*, 1260901–1260901.

(2) Biffi, S.; Cerbino, R.; Bomboi, F.; Paraboschi, E. M.; Asselta, R.; Sciortino, F.; Bellini, T. Phase Behavior and Critical Activated Dynamics of Limited-Valence DNA Nanostars. *Proc. Natl. Acad. Sci.* **2013**, *110*, 15633–15637.

(3) Um, S. H.; Lee, J. B.; Park, N.; Kwon, S. Y.; Umbach, C. C.; Luo, D. Enzyme-Catalysed Assembly of DNA Hydrogel. *Nat. Mater.* **2006**, *5*, 797–801.

(4) Davis, J. T.; Spada, G. P. Supramolecular Architectures Generated by Self-Assembly of Guanosine Derivatives. *Chem. Soc. Rev.* **2007**, *36*, 296–313.

(5) Cafferty, B. J.; Gallego, 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.

(6) 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.

(7) 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.

(8) Bellini, T.; Zanchetta, G.; Fraccia, T. P.; Cerbino, R.; Tsai, E.; Smith, G. P.; Moran, M. J.; Walba, D. M.; Clark, N. A. Liquid Crystal Self-Assembly of Random-Sequence DNA Oligomers. *Proceedings of the National Academy of Sciences*, 2012, *109*, 1110–1115.

(9) Selinger, J. V.; Bruinsma, R. F. Hexagonal and Nematic Phases of Chains. I. Correlation Functions. *Phys. Rev. A.* **1991**, *43*, 2910–2921.

(10) Cates, M. E.; Candau, S. J. Statics and Dynamics of Worm-like Surfactant Micelles. *J.*

Phys. Condens. Matter **1999**, *2*, 6869–6892.

(11) SantaLucia, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.

(12) Peyret, N. Prediction of Nucleic Acid Hybridization: Parameters and Algorithms, Wayne State University, Detroit, 2000.

(13) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-Stacking and Base-Pairing Contributions into Thermal Stability of the DNA Double Helix. *Nucleic Acids Res.* **2006**, *34*, 564–574.

(14) 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*, 10366–10383.

(15) 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.

(16) Mitchell, J.; Tiddy, G. J. T.; Waring, L.; Bostock, T.; McDonald, M. P. Phase Behaviour of Polyoxyethylene Surfactants with Water. *J. Chem. Soc., Faraday Trans.* **1983**, *79*, 975–1000.

(17) Rastogi, R. P. Thermodynamics of Phase Equilibria and Phase Diagrams. *J. Chem. Educ.* **1964**, *41*, 443.

(18) Kolli, H. B.; Frezza, E.; Cinacchi, G.; Ferrarini, A.; Giacometti, A.; Hudson, T. S.; De Michele, C.; Sciortino, F. Self-Assembly of Hard Helices: A Rich and Unconventional Polymorphism. *Soft Matter* **2014**, *10*, 8171–8187.

(19) 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.

(20) 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.

(21) 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.

(22) Brandes, R.; Kearns, D. R. Magnetic Ordering of DNA Liquid Crystals. *Biochemistry* **1986**, *25*, 5890–5895.

(23) Oldenbourg, R.; Ruiz, T. Birefringence of Macromolecules. Wiener's Theory Revisited, with Applications to DNA and Tobacco Mosaic Virus. *Biophys. J.* **1989**, *56*, 195–205.

(24) Biffi, S.; Cerbino, R.; Nava, G.; Bomboi, F.; Sciortino, F.; Bellini, T. Equilibrium Gels of Low-Valence DNA Nanostars: A Colloidal Model for Strong Glass Formers. *Soft Matter* **2015**, *11*, 3132–3138.

(25) Mariani, P.; Spinozzi, F.; Federiconi, F.; Amenitsch, H.; Spindler, L.; Drevensek-olenik, I. Small Angle X-Ray Scattering Analysis of Deoxyguanosine 5′-Monophosphate Self-Assembing in Solution : Nucleation and Growth of G-Quadruplexes. *J. Phys. Chem. B*

2009, *113*, 7934–7944.

(26) 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.

(27) Horn, T.; Urdea, M. S. A Chemical 5′-Phosphorylation of Oligodeoxyribonucleotides That Can Be Monitored by Trityl Cation Release. *Tetrahedron Lett.* **1986**, *27*, 4705–4708.