Macromolecules

pubs.acs.org/Macromolecules Article

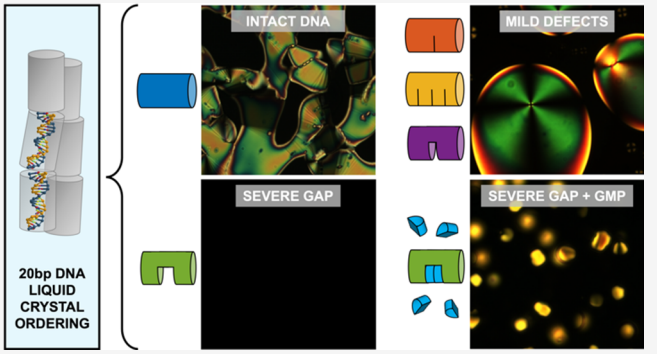
Liquid Crystal Ordering in DNA Double Helices with Backbone Discontinuities

Francesco Fontana, Tommaso Bellini, and Marco Todisco*

Cite This: https://doi.org/10.1021/acs.macromol.2c00856 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Short DNA double helices in highly concentrated solutions self-assemble into long noncovalent polymers and globally order as liquid crystals. What level of helical defect is compatible with this collective ordering is still unclear. In this work, we show that a series of 20 bp-long DNA double helices with different structural defects such as nicks and gaps still retain their capability to transition into liquid crystalline solutions despite the increased flexibility of their structure. These results expand our understanding of the space of hybridization motifs leading to linear physical polymerization and liquid crystal ordering. Moreover, by studying liquid-crystalline solutions made of 20 bp-long duplexes with a central dinucleotidic gap, we find that regardless of the highly crowded state of the system, guanosine monophosphate molecules can diffuse in-between the DNA polymers and effectively fill the gap. This finding enables formulation of a prebiotic scenario where fragmented DNA double helices, held together by self-interactions and supramolecular ordering, form a matrix rich in substrates for ligation, gap-filling, and primer extension reactions.



INTRODUCTION

Hydrogen bonding and hydrophobic base stacking are the fundamental interactions governing DNA base pairing and leading to the formation of double helices and to the propagation of genetic information. Aside from their well-recognized major role in stabilizing the helical structure,[1] attractive stacking interactions between distinct duplexes with blunt-ended terminals are significant[2,3] and are known to drive the formation of noncovalent polymers in concentrated solutions of short double helices.[4] Specifically, upon increasing the concentration of well-paired oligomers in aqueous solution, the system undergoes a sharp transition from an isotropic state—formed by isolated duplexes or small duplex aggregates—to a liquid crystalline (LC) state in which long linear chains of duplexes are coherently aligned.[4−6] Instrumental to LC ordering are the stacking interactions between blunt-ended duplex terminals and the duplex linearity and stiffness, with bent shapes and high flexibility known to destabilize the collective ordering and linear aggregation.[7−9] However, after their discovery,[4] it was found that the formation of LC phases in solution of oligonucleotides was way more ubiquitous than expected, showing up both in DNA and RNA,[10] in the system composed of very short strands hybridizing only in the LC phase[11,12] and in various solutions of DNA oligomers with fully or partially random sequences,[13] a condition where only small fractions of double helices are expected to be perfectly pairing and driving the phase transition.[14]

These observations raise the question of how extended it is the sequence space in which LC ordering appears. Previous investigations have studied the relationship between duplex termini and end-to-end interaction strength.[13,15,16] Here, we focus instead on flexibility and bending of duplex cores by altering the properties of the mesogenic unit through discontinuities along the DNA strands (for simplicity, we will refer to these as defects) and studying their effects on the LC phase diagram. Specifically, we present a careful characterization of a series of 20 bp-long blunt-ended double helices designed to include a single nick (NCK), three nicks (3NCK), a single nucleotide gap (GAP), or a dinucleotidic gap (2GAP). Comparing the extension of the LC ordering among these species and their intact counterpart (INT) as a function of concentration (c) and temperature (T) enables the evaluation of their relative stability and to better understand the requirements to produce a stable LC solution.

In this study, we also investigate the behavior of more heterogeneous systems. In particular, we exploit the LC

Received: April 25, 2022
Revised: June 15, 2022

ordering and supramolecular assembly of 2GAP to explore whether free guanosine monophosphate (GMP) can bind to the structure by noncovalently filling its −CC− gap, thus opening the possibility for a series of relevant chemical reactions that can happen where a defect is present, such as covalent gap-filling,[17] primer extension,[18] and on-template[19] or off-template ligation.[20−22]

MATERIALS AND METHODS

Stock Solution Preparation. PEG8k stock solutions are prepared by resuspending the powder (Fluka) in MilliQ water at different concentrations ranging from 150 to 500 g/L. GMP stock solutions are prepared by resuspending the GMP sodium salt powder (Sigma Aldrich) in MilliQ water to reach a concentration of 17.5 mM. Every DNA oligonucleotide used in this work has been acquired from Integrated DNA Technologies (IDT). Since oligomers are used at extremely high concentrations, they must first be cleaned up from residual salt and synthesis byproducts. The lyophilized powder from the producer is first diluted in MilliQ water at 20 g/L and centrifuged at 13 k rpm (Eppendorf 5418 microcentrifuge) for 1 h to pellet any solid particles that might be present in the suspension. The supernatant of the centrifuged solution is then dialyzed overnight against 2 L of 60 mM sodium chloride (powder acquired from Sigma Aldrich) solution and then lyophilized. The material obtained this way is weighed and resuspended in MilliQ water at 50 g/L. The DNA concentration is determined through UV spectrophotometry (Thermo Scientific Evolution 350 UV−vis spectrophotometer) using extinction coefficients reported by IDT. Once the purity (DNA/total material) is checked to be consistent with a low-salt and high-purity solution, a stock solution is prepared by diluting DNA to reach a final concentration of 7 mM.

Microscopy Sample Preparation. The DNA samples for microscope observations are prepared by pipetting multiple drops of DNA stock solutions on a glass slide. The water content of each drop is evaporated in air while keeping the glass slide on a plate heater at 40 °C to improve dryness, and new drops are added until the desired target concentration is reached. Target concentration in g/L for a DNA solution as a function of added drops of stock solution can be obtained through the following simple formula:

$$c_{\mathrm{DNA}} = \frac{V_{\mathrm{drop}} \cdot n \cdot \sum_i c_i}{V_i + \frac{V_{\mathrm{drop}} \cdot n}{\rho_{\mathrm{DNA}}} \sum_i c_i} \quad (1)$$

where c_i is the concentration of the stock solution in g/L, n is the number of drops added, and V_{drop} is the volume of each drop, which is typically set at 0.5 μL. When DNA is mixed with PEG and/or GMP, the solution is resuspended in a final drop of PEG or PEG + GMP. After the last drop addition, we reach the target concentration and cover the sample by placing a second glass slide on top of the solution, assembling a so-called microscope cell. Then, 20 μm silica spacer rods are placed all around the sample in between the slides in order to keep them separated at a fixed distance for better reproducibility. The samples sandwiched in the microscope cells obtained this way are isolated from surrounding air using y25 fluorinated oil (Fomblin) in order to maintain stability over time and prevent evaporation at high temperatures during thermal cycles. Before making any observation on the samples, they are equilibrated at high temperatures (≃60 °C) until homogenization is reached and then slowly cooled down to room temperature. To probe the equilibrium of the sample, we use visual clues: if the LC phase fills the whole volume and the textures are homogeneous and melt at the same temperature in every point in space, then we can confidently say that the sample reached equilibrium. If this is not the case, we repeat the equilibration process by reheating it.

Microscopy Observations. Sample observation is performed using a Nikon TE200 optical microscope with crossed polarizers in order to identify the typical birefringent textures of LC domains. Temperature of the microscope stage is controlled using an Instec mK2000 Peltier module. All micrographs were acquired using a Nikon DS-Fi3 camera.

UV−Vis Melting Experiments. Melting temperature of DNA double helices in the diluted regime is measured through UV−vis melting using a Thermo Scientific Evolution 350 UV−vis spectrophotometer. We acquire the absorbance spectra by increasing the temperature between each cycle ($T_i \simeq 20$ °C, $T_f \simeq 90$ °C, and rate = 1 °C/cycle), and the absorbance at 260 nm as a function of temperature is processed. Data are corrected for lower and upper baselines, and normalized signals provide a readout of the melting temperature of the double helix at an signal amplitude equal to 0.5.

Fluorescence Melting Experiments. Melting temperature of DNA double helices at high concentrations cannot be easily measured using a conventional spectrophotometer due to signal saturation. In order to perform such measurements, we exploited the properties of GelRed (Biotium) whose quantum yield increases when in contact with double helices. DNA samples in microscopy cells are illuminated at 550 nm using a LED source (CoolLED pE-4000), and fluorescence of the sample at 630 nm is captured using a Nikon DS-Fi3 camera. Average fluorescence as a function of T is then corrected for lower and upper baselines, and its normalized signal is used to determine the melting temperature at an signal amplitude equal to 0.5.

Coarse-Grained Molecular Dynamics Simulations. To investigate the structural properties of studied duplexes, we performed a series of molecular dynamics (MD) simulations using oxDNA.[23−25] Starting configurations were generated using Tiamat[26] and converted to the oxDNA format using TacoxDNA.[27] Each duplex initial configuration was minimized and relaxed as recommended[28] with default mutual traps produced by TacoxDNA. The system was then simulated without additional forces for 10^8 steps with dt = 0.003 using the DNA2 model, salt concentration = 1, john thermostat, Newtonian steps = 10^3, pt = 0.1, diffusion coefficient = 2.5, and sequence-specific interactions.[24] A configuration was exported every 10^4 steps, converted to the *xyz* format, and imported in MATLAB for analysis. To study the structural conformation of the molecules, the 3′ half backbone elements of the template strand were aligned using Horn's method[29] to a reference B-DNA structure produced with Tiamat, and the median structure was calculated as a reference for comparisons. To study the molecule bending, the angle between the two halves of the molecules is computed. The vectors used to determine the angle are calculated between the 1st and 10th-backbone element of the template strand and the 11th and 20th elements. A histogram is calculated to capture the flexibility and preferential angle assumed by the duplex. Fraying of nucleobases is studied by analyzing the distance between interacting bases as a function of time. The threshold for the hydrogen bonds is set at 0.6 simulation units, so that every time the two nucleobases adjacent to a defect diffuse at a distance larger than 0.6 simulation units, we annotate a fraying event. These events are summed up among all flaking sites and binned with a 0.5 ns time window for plotting.

RESULTS AND DISCUSSION

Phase Diagrams and Melting of DNA Double Helices with Structural Defects. In this work, we aim at understanding how alterations in the central part of double helical structures affect the formation of DNA-based LC phases. We have designed a series of 20 bp DNA double helices with equal blunt-ended termini as sketched in Figure 1a as intact (INT) or having either a single nick (NCK), three nicks (3NCK), a single nucleotide gap (GAP), or a dinucleotidic gap (2GAP). All sequences share the same common 20 nt-long template strand. For each of these species, we prepare a series of microscopy cells with DNA samples at various concentrations (c_{DNA}) up to 400 g/L (≈60 mM), and for each sample, polarized optical microscopy observations are performed at different T to evaluate the presence of LC ordering. The map of these observations in the $c_{\mathrm{DNA}}-T$ plane identifies a phase diagram as depicted in Figure 1b, where open and closed dots

https://doi.org/10.1021/acs.macromol.2c00856
Macromolecules XXXX, XXX, XXX−XXX

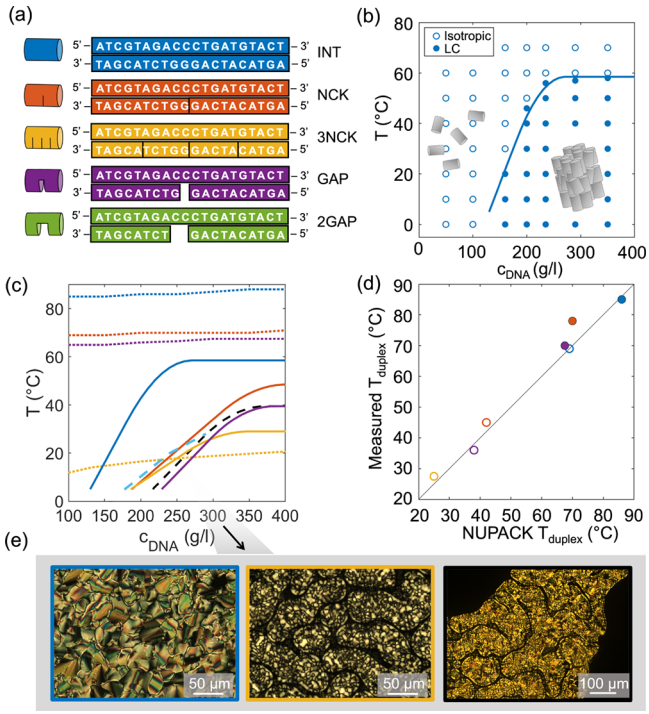


Figure 1. (a) Sketches of double helices used in this work, with the sequences used to form them and their identifying name and color. (b) Phase diagram (c_{DNA} vs T) of 20 bp intact dsDNA (INT) suspended in aqueous solution. The phase diagram is built by observing samples of different c_{DNA} through optical microscopy with crossed polarizers. We cool the samples from high T and take note of the appearance of birefringence signaling the formation of liquid crystal ordering. The solid line marks the boundary between the isotropic phase (empty circles) and liquid crystal ordering (filled circles). (c) Full lines: phase boundaries of each system obtained through polarized optical microscopy. Dotted lines: NUPACK-predicted T_{duplex} of the various double helices. Black dashed line: phase boundary of a self-complementary decamer (10S). Blue dashed line: phase boundary of a mixture of INT (80%) and 2GAP (20%). (d) Correlation between NUPACK predictions and experimental data for T_{duplex} of studied double helices. Data are obtained through UV spectrophotometry (empty circles) and fluorescence microscopy (filled circles). (e) Liquid crystal phases at c_{DNA} = 300 g/L as observed in polarized optical microscopy. Textures of LC formed by defected helices, such as 3NCK (yellow frame), appear different from those of the INT system (blue frame) and similar to those formed by the self-complementary decamers (black frame), NCK and GAP (not shown).

mark conditions in which the system is in isotropic and LC phases and where the phase boundary (line) is obtained by data interpolation. The area of the LC phase, i.e., its extension toward large T and small c_{DNA}, is an indication of the stability of LC supramolecular assembly. As defects are introduced in the double helices, the phase boundaries retain the same basic features while being shifted to higher c_{DNA} and lower T (Figure 1c). Comparison between the different phase boundaries reveals a hierarchy of stability: at low T, by ordering the values of c_{DNA} where the transition to LC occurs, we find that INT > NCK > 3NCK > GAP >2GAP. The latter is so heavily destabilized by its defect that it does not produce any detectable phase at least up to 400 g/L and its phase boundary is thus not shown here. Upon increasing T, the hierarchy changes in a nontrivial way, with 3NCK phases disappearing at lower T and crossing the GAP phase boundary. Figure 1c also shows, for comparison, the phase boundary of a solution of a self-complementary DNA 10mer (10S, 5′-CGACTAGTCG-3′, black dashed line). Such a phase diagram lies close to those of NCK and GAP, which share with 10S the length of the well-formed sections of double helices, and these three together with 3NCK converge to a similar concentration (≈200 g/L) at low temperatures, suggesting that their structures diverge as T increases due to the different nature of their structural defects.

Since the LC phase depends on the presence of stable double helices in solution, we study to what extent the observed loss of stability of the supramolecular assemblies simply reflects a difference in the melting T of the duplexes (T_{duplex}). We perform a set of melting experiments at low ($c_{DNA} \approx 5\ \mu M$) and high concentrations (up to $c_{DNA} \approx 50$ mM), whose results are shown in Figure 1d scattered against the values predicted by the software NUPACK[30] using the nearest neighbor (NN) model for nucleic acid thermodynamics to determine the energy of multistrand structures. We find that the agreement is good even at high concentrations, even though the NN parameters were optimized in a diluted regime. Since we have validated NUPACK against our defected duplexes, we use its predictions to extract the c_{DNA} dependence of T_{duplex} over the whole range of concentrations used in this work, as shown in Figure 1c (dotted lines), where it is possible to appreciate some level of correlation between T_{duplex} and T_{phase} (the melting temperature of the LC phases). Interestingly, we find in the case of 3NCK that T_{phase} lies extremely closely to T_{duplex}, pointing to a low stability of its supramolecular ordering at high temperatures as a direct consequence of 3NCK duplex melting. In all the other cases, we find that ($T_{duplex} - T_{phase}$) > 20 °C (Figure 1c, compare continuous and dotted lines), indicating that the LC destabilization in such cases is not due to the melting of the duplexes but rather lies in other features of the supramolecular assembly, which must be investigated more in depth.

Coarse-Grained MD Simulations. Since the shape of the DNA double helices is known to affect the overall supramolecular assembly,[7−9] we study here the conformations accessible by our molecules, performing a series of coarse-grained MD simulations with oxDNA, a nucleotide-level model optimized to describe the thermodynamics of hybridization and the structure and mechanical properties of single-stranded and duplex DNA.[23−25]

We run a series of simulations and extract the median conformation acquired by the molecules as representative for their structure, shown here for a direct comparison (Figure 2a). We can see that even though the supramolecular ordering of NCK and 3NCK is significantly destabilized with respect to INT, only minor differences can be found among structures assumed by their double helices. This result agrees with previous studies on nicked molecules[31] and suggests that these subtle differences are amplified in the process of producing LC ordering.

Simulations performed on GAP molecules instead show that this molecule has a markedly distinct behavior with respect to INT, NCK, and 3NCK. Literature studies in double helices with a single gap have shown that these are not isotropically flexible around their defect. Instead, they can maintain a B-DNA structure while preferentially bending in a defined direction.[31,32] This phenomenon is correctly modeled by oxDNA, with GAP preferentially acquiring a more bent structure and showing an increased flexibility. To capture this feature, we show in Figure 2b the angle θ measured

https://doi.org/10.1021/acs.macromol.2c00856
Macromolecules XXXX, XXX, XXX−XXX

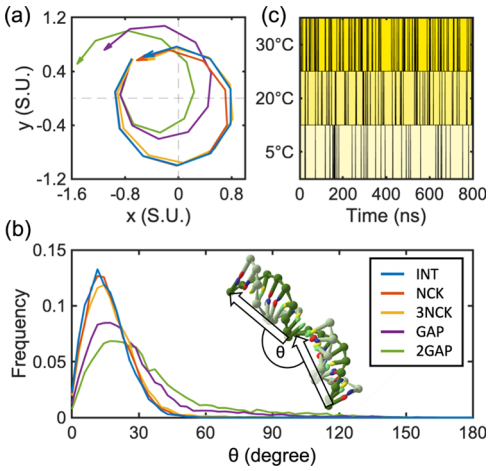


Figure 2. (a) Projection in the *XY* plane of the 5′ half of the 20 nt template strand for studied duplexes at 20 °C (color code in panel b). Each structure is obtained by aligning via roto-traslation the 3′ half of all double helices over a reference B-DNA structure. The median structure of the 5′ half of the template strand is extracted and depicted here as representative of the backbone spatial configuration. The arrow points to the 5′ end of the strand. (b) Angle θ between the two halves of the template strand at 20 °C. The vectors used for the angle calculation are defined as sketched in the 2GAP duplex snapshot. As gaps are introduced in the duplex, it becomes more flexible and preferentially bent. (c) Fraying time-courses in 3NCK at 5, 20, and 30 °C. Each time series is obtained by 0.5 ns binning of fraying events at sites flanking nicks (nucleobases in positions 5, 6, 10, 11, 15, and 16 on the template strand). Every vertical black line corresponds to the presence of at least one fraying event among all monitored positions in the corresponding 0.5 ns time window. Fraying events become more frequent as the temperature increases.

between the two halves of the duplexes, which shows both peak shift (change in preferential θ) and a severe spreading (enhanced flexibility) for gapped molecules.

Simulations of 2GAP show a deformation qualitatively similar but more severe than GAP. The double gap makes this molecule more flexible around its defect, enabling accessing extremely bent conformations with the two double helix halves freely rotating while connected by the backbone behaving as a flexible linker.

Aside from shedding light on structural features, oxDNA can be used to study fraying on pairs of terminal nucleobases,[33] a phenomenon with a significant impact on coaxial stacking, which is known to be suppressed by dangling ends.[21] Fraying, which consists in the transient opening and closing of the terminal nucleotides of double helices, is observed in our MD simulation and summarized in Figure 2c for the study of 3NCK. During the simulated experiment, we annotate whether at least one fraying event occurs every 0.5 ns—shown here as a vertical black line—accounting for the transient opening and closing of the upstream or downstream nucleobases immediately adjacent to the three different nicks present in the double helix. This study has been repeated at three different T (5, 20, and 30 °C) to gather a clearer understanding of the phenomenon in the context of 3NCK.

Liquid-Crystalline Phase Stability of Nicked and Gapped Double Helices. The LC-forming units in solutions of NCK, GAP, 2GAP, and 10S can be envisioned as made by two 10 bp-long duplexes (10 bp + 9 bp for GAP and 10 bp + 8 bp for 2GAP) held together by stacking forces alone (10S) or combined with the proximity constraint due to a continuous backbone (NCK) or a backbone with 1 or 2 unpaired bases (GAP and 2GAP). The phase behavior of NCK clearly indicates that the linear stacking of the two halves of the LC-forming unit is favored by the continuity of the template backbone. The same cannot be said for the GAP LC-forming unit. Indeed, the nearly identical phase diagrams of GAP and 10S indicate that the entropic advantage brought about by the chemical continuity of the GAP template, by which the two halves are always in contact, is compensated by the more difficult packing and collective ordering resulting from the bent GAP structure (Figure 2b). The similarity of GAP and 10S is not limited to the LC phase boundaries but also to the specific molecular arrangement: at 300 g/L, both GAP and 10S are in the chiral nematic phase, a form of LC ordering less regular than the columnar phase ordering observed at the same concentration in INT solutions. The two LC phases can be distinguished by observing their textures in polarized optical microscopy (Figure 1e).

A similar yet more extreme behavior is observed in 2GAP. MD simulations indicate that 2GAP has a larger mean internal bend than GAP and that it can more easily form highly bent structures, as indicated by the larger tail of the angular distribution in Figure 2b. Phase behavior study indicates that the two-base-long gap makes the stacking between the two helical halves of 2GAP too weak and the structure too bent to form LC order. These observations indicate that a flexible linker between the double helices is not neutral at all for their stacking, but rather it destabilizes both the molecular structure and the supramolecular ordering.

The relevance of bending and flexibility in the phase behavior of GAP and 2GAP suggests that these must be the factors that reduce the stability of the LC phase in NCK and 3NCK with respect to the INT duplex too. Indeed, at low T, the LC phase of NCK and 3NCK is found at a concentration of about 50 g/L larger than in INT, an effect that has to be due to the extra flexibility of these helices given by the nicks. According to the well-established literature, it is possible to determine the fraction of unstacked nicked duplexes at any time, an evaluation experimentally confirmed by the observable effects of one nick on the migration of a double helix in a polyacrylamide gel.[1] We can hypothesize that the destabilization measured in NCK is all due to a small subpopulation of duplexes where the nucleobases flanking every nick are unstacked, rendering the adjacent helice stretches free to rotate in conformations—according to that seen in GAP and 2GAP—that will impair LC formation in the explored c_{DNA} range. To test this hypothesis, we prepared mixtures of INT and 2GAP to reproduce a solution composed of a rigid population (INT) and a flexible population (2GAP), emulating what we would expect in NCK solutions. Indeed, we find that a mixture made of 20% flexible molecules is enough to heavily impact the phase behavior of INT and effectively reproduce the phase boundary of NCK in a range studied here from 0 to 30 °C (light blue dashed line in Figure 3c). This piece of data provides a simple understanding of the nicked duplex behavior as mediated by its unstacked state: following this model, we can treat that the NCK mixtures are composed at any time of 80% stacked NCK molecules and 20% unstacked NCK molecules, an equilibrium population corresponding to a free energy difference of ∼1.0 kcal/mol, a value in agreement with literature data on stacking energy.

Although these notions provide a satisfactory explanation for the observed data, they need to be reconciled with the results

https://doi.org/10.1021/acs.macromol.2c00856
Macromolecules XXXX, XXX, XXX−XXX

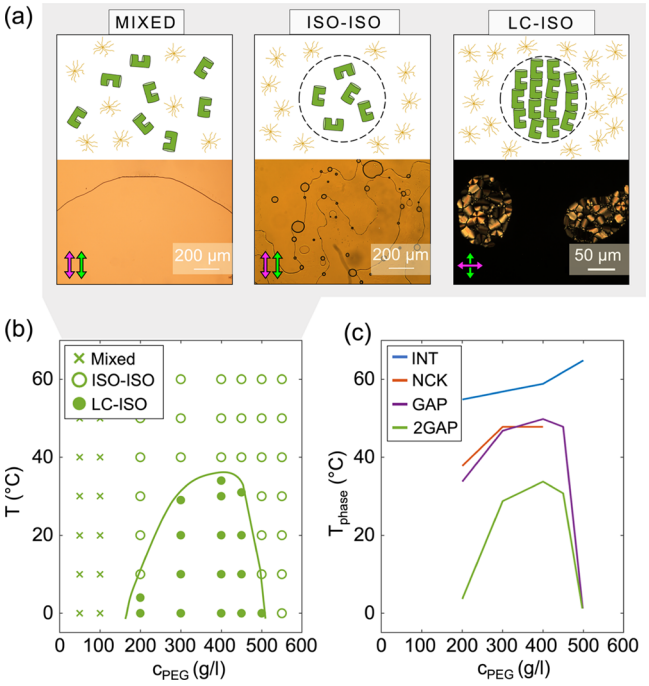


Figure 3. (a) Sketches (top) and microscope pictures (bottom) illustrating the behavior of DNA−PEG mixtures. The samples can be either mixed (left panel) or produce a phase separation in which a PEG-rich solution hosts DNA-rich domains either in the isotropic state (middle panel) or with LC ordering (right panel). Purple and green arrows show the relative orientation of the microscope analyzer and polarizer setup for image acquisition. (b) Phase diagram (c_{PEG} vs T) of 2GAP double helices suspended in PEG solutions. The solid line marks the boundary between the isotropic phase (empty dots) and liquid crystal ordering (full dots) in the phase-separated DNA-rich domains, as recognized by comparing images taken with crossed and uncrossed polarizers. At low c_{PEG}, the system remains mixed and homogeneous (crosses). (c) Phase diagrams (c_{PEG} vs T) of the various short dsDNA species in PEG solutions. We highlight the boundary for each species, as previously defined, using solid lines.

of the MD simulations that do not show significant differences between the double helical structures of INT and NCK in diluted conditions. Indeed, oxDNA results suggest that the timescale for stacking/unstacking of nicked duplexes is too short for the two halves to completely decorrelate their orientation. The origin of the contradiction might be in the effects of crowding. We argue that, in analogy to the collapse of flexible polymers induced by crowded environments,[34] nicked helices might more easily fold because of the interactions with either the polyacrylamide gel network[1] or with the surrounding copies of the same duplexes as seen in this work.

The effect of nicks is also evident at large c_{DNA} in the thermal stability of the LC phase, which is markedly reduced with respect to INT. We understand this effect as a combination of the decrease of the stacking free energy with T and of fraying, which transiently disrupts stacking at the nicks. Indeed, our study of 3NCK fraying (Figure 2c) indicates a marked increase in the frequency of fraying events as T rises, an effect that necessarily promotes local flexibility and further destabilizes supramolecular ordering.

Liquid-Crystalline Phases in Solutions of DNA Double Helices and PEG. Since it is problematic to produce homogeneous solutions of DNA at concentrations higher than 400 g/L, we rely here on the properties of PEG to produce highly concentrated DNA-rich domains via phase separation whose concentration is governed by the osmotic pressure provided by PEG.[35,36] This strategy enables exploring large values of c_{DNA} and hence determines whether the absence of LC phases of 2GAP was a limitation due to concentrations or if it is instead caused by some intrinsic impairment in its supramolecular assemblies. We prepare a series of microscopy flat cells with DNA−PEG mixtures and study their behavior as a function of PEG concentration (c_{PEG}) and temperature using polarized optical microscopy. At low PEG concentrations, the solution remains uniform—no phase separation—with DNA and PEG homogeneously mixed. Upon increasing c_{PEG}, the solution undergoes a phase separation, where small DNA-rich domains appear with local c_{DNA} large enough to induce DNA LC ordering and residual PEG concentration in the order of only a few g/L.[36] As the T increases above T_{phase}, the LC phase inside the DNA-rich domains becomes isotropic (Figure 3a). From these observations, we can once again assemble phase diagrams, whose phase boundaries in the $c_{PEG}-T$ space convey information on the relative stability of the different species studied in this work (Figure 3b). Using this approach, we confirm that NCK and GAP have a very similar behavior and that they are both less stable than INT. Moreover, we can observe liquid crystal ordering of 2GAP, confirming that the supramolecular ordering of this species is dramatically destabilized by the presence of the double gap but not completely impaired.

The comparison between the phase diagrams in Figures 1c and 3c enables us to obtain a rough estimate of the DNA concentration inside the DNA-rich domains in PEG using T_{phase} as the reference: for c_{PEG} = 150, 200, 250, and 300 g/L, we can determine c_{DNA} of about ≈225, 310, 360, and 410 g/L, respectively, starting from an average c_{DNA} equal to 46 g/L, which indicates that phase separation and PEG osmotic pressure effectively enhance the DNA concentration with respect to the mixed state by factors of 4.9, 6.7, 7.8, and 8.9, respectively.

GMP Increases 2GAP Phase Stability. Since the significant destabilization of 2GAP supramolecular ordering is due to the absence of two consecutive nucleotides and the resulting increase in the molecule flexibility, we investigate whether it is possible to partially restore the molecular rigidity by filling the −CC− gap using GMP. In other words, we investigate if the molecular environment around the gap can act as a binding site for GMP and exploit the phase behavior as an indicator of such binding.

We prepare a set of microscopy cells with DNA−PEG by adding an excess of GMP ([GMP] = 10 · [DNA]) in a large enough amount to produce effects on LC phases that are easily measurable without causing any precipitation or apparent anomaly. We focus our attention on the phase behavior of NCK, GAP, and 2GAP and compare it with the one obtained without GMP, finding that the addiction of GMP has nontrivial effects depending on the DNA oligonucleotides used in the system, having either a positive, neutral, or negative influence on supramolecular ordering. The addiction of GMP to NCK−PEG mixtures causes a reduction of LC-phase stability (see Figure 4a, red symbols vs red line), an effect already observed when extra molecules are solubilized in the crowded environment of DNA LC: when these additional solutes have sizes comparable with the LC interstitial space, their presence hampers self-assembly and ordering.[21] We can speculate here that this effect may be raising from (i) GMP

https://doi.org/10.1021/acs.macromol.2c00856
Macromolecules XXXX, XXX, XXX−XXX

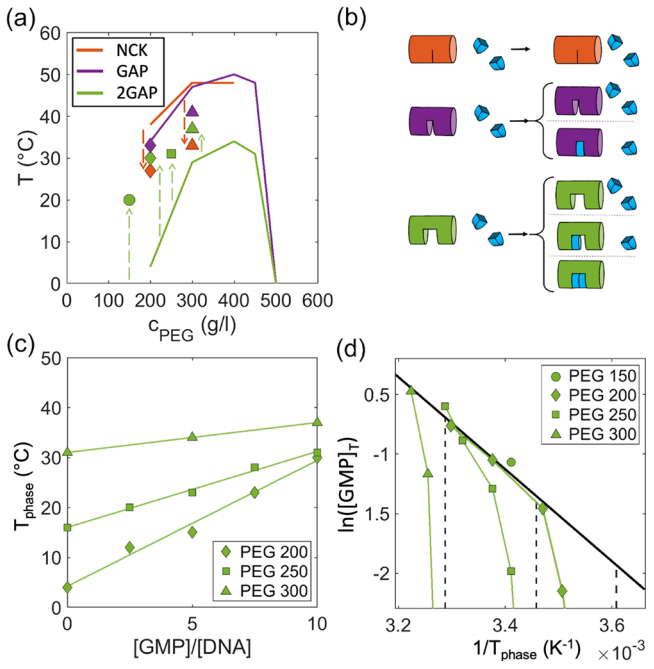


Figure 4. (a) Result of the solubilization of guanosine monophosphate (GMP) in DNA/PEG solutions. Solid symbols represent LC boundaries in samples with [GMP] = 10·[DNA]. The color code is given in panel b. Phase boundaries without GMP (from Figure 3c) are also reported to provide a reference. (b) Sketches of GMP−DNA interaction at defected sites: no binding is expected for NCK, while a gap filling process is possible for both GAP and 2GAP. (c) Progressive stabilization of 2GAP−PEG LC phases due to GMP. We plot the liquid crystal melting temperatures (T_{phase}) as a function of GMP molar concentration for three different concentrations of PEG. (d) van't Hoff plot for 2GAP and GMP binding as defined in the main text. All datasets align on a straight line for large $[GMP]_T$, giving insight on the binding strength. For low $[GMP]_T$, the phase stability is no longer determined by GMP binding and therefore diverges to the melting temperature of the 2GAP−PEG phase (dashed lines).

directly interfering with LC ordering, for example, by stacking on DNA duplex termini and competing with end-to-end aggregation or (ii) by GMP altering the phase separation by diluting the DNA-rich phase, for example, producing aggregates[37,38] that cannot effectively fit in the LC interstitial spaces and that would result in an increase of the DNA-rich domain volume. We could not observe any significant change in the phase-separated volume as GMP is added, supporting either the hypothesis of a direct interaction or that the dilution effect is so subtle that polarized microscopy observation can hardly capture it.

GMP has almost no effect on the LC ordering of GAP (purple symbols vs purple line), resulting in LC phases that are more stable than those of NCK−PEG−GMP. Since the phase diagram of NCK and GAP is very similar both with or without PEG in solution, the presence of GMP in the interstitial spaces between columns of aggregated GAP duplexes should destabilize the ordering as it does for NCK or even more, given the larger flexibility of GAP. The minimal destabilization observed here suggests that GMP, to some extent, interacts with GAP and restores the helical continuity. The binding of GMP in the double helical gaps is much more evident in the behavior of 2GAP, whose LC ordering is greatly stabilized by GMP (green symbols vs green line), with its thermal stability larger than that in NCK−PEG−GMP. Remarkably, the addition of GMP favors the appearance of LC ordering in conditions where it is otherwise not observed, such as for c_{PEG} = 150 g/L. The 2GAP stabilization provided by GMP is progressive, as we can observe by preparing three sets of 2GAP−PEG−GMP microscope cells with c_{PEG} = 200, 250, 300 g/L and adding GMP at different initial concentrations ranging from 0 mM up to 70 mM (corresponding to [GMP] = 10 · [DNA]). We observe that the 2GAP supramolecular order is increasingly stabilized by GMP as a function of its concentration (Figure 3a).

We interpret this enhanced LC phase stabilization of 2GAP−PEG as the result of the binding of GMP to 2GAP. Since T_{phase} of 2GAP−PEG−GMP is always lower than T_{phase} of GAP−PEG, it is reasonable to assume that it is controlled by GMP binding. In other words, the 2GAP−PEG−GMP LC phase melts whenever the temperature is too large for GMP to remain bound to 2GAP, which is analogous as saying that the concentration of GMP in the system equals its binding constant.

Accordingly, the melting temperature can be expressed in terms of the enthalpy (ΔH) and entropy difference (ΔS) between the bound and unbound states using the van't Hoff equation:

$$T_{phase} = \frac{\Delta H}{R \cdot \ln([GMP]_T) + \Delta S} \quad (2)$$

where $[GMP]_T$ is the total concentration of GMP that determines the experimentally observed T_{phase}, equal to the dissociation constant of GMP binding to the dinucleotidic gap. Since the basic assumption in eq 2 is that the LC melting is GMP binding-limited, we expect this equation to hold for large enough [GMP]/[DNA], where the behavior of the solution does not depend strongly on the phase behavior of 2GAP by itself. We do not expect it to hold for small values of [GMP]/[DNA], since the LC phase in this regime has instead a GMP-independent melting temperature. To reach a quantitative estimate from eq 2, we assume that GMP in PEG is phase separated with DNA, and thus, its concentration is enhanced by the same factor as c_{DNA}.

To proceed with our analysis, we extract ΔH and ΔS from the van't Hoff plot of T_{phase} and $[GMP]_T$ (Figure 4d). In such a plot, we expect our data to align on a straight line if the measured binding constants belong to a single reaction with enthalpy and entropy changes that are temperature-independent. In our case, we can see how our data points lie on a straight line (black continuous line) for [GMP]/[DNA] > 2, in line with what was expected. As [GMP]/[DNA] becomes low, instead, our datasets diverge toward the melting temperature of the 2GAP−PEG mixtures without any GMP (black dashed lines, each one for a different PEG concentration). From the slope and intercept of the line for the GMP-limited melting, we can determine best fitting ΔH equal to −8.1 kcal/mol and best fitting ΔS = −25 cal/mol/K. These values are comparable with literature data on GMP binding on a DNA template in isotropic conditions[39] and support our description of the phase stabilization through direct GMP binding. It can be noted that any diluting effects on DNA and/or GMP, such as the one potentially coming from GMP aggregation, would make our calculated values a lower estimate for GMP binding strength to a double gap in a LC phase.

https://doi.org/10.1021/acs.macromol.2c00856
Macromolecules XXXX, XXX, XXX−XXX

CONCLUSIONS

In this work, we have studied the LC phases that form, via collective supramolecular ordering, in solutions of a set of oligonucleotides chosen to share a large part of their sequence but differ for the presence of structural defects of various kinds. Our findings are on three levels.

Through a systematic study of phase diagrams (concentration vs temperature) of species having an intact backbone or discontinuities such as a single nick, multiple nicks, one gap, or two gaps, we found that supramolecular ordering in DNA is highly resilient to duplex alterations to a degree not recognized before. We indeed observed LC ordering in all species we considered, provided that the appropriate concentrations and temperatures are met.

The collective ordering appears as a powerful tool to appreciate the structural differences between nucleic acid duplexes, in particular bending, flexibility and duplex melting. A clean hierarchy of phase stability can be recognized, which reflects the distinct structural features of the species under investigation. The observations agree with literature structural studies and coarse-grained MD simulations presented here. A noteworthy example is the case of 3NCK whose phase boundary crosses the one of GAP at increasing temperatures, indicating a better formed structure of 3NCK at low T and a larger thermal stability of GAP.

By exploiting the phase behavior as a gauge for the duplex structure, we have shown that 2GAP can assemble into LC phases when mixed with PEG8k, and its behavior can be tuned by doping GMP in a concentration-dependent fashion. We understand and study this phenomenon as due to a gap-filling process whose association energy difference, here extracted from the LC phase boundary, we find to be comparable with GMP binding energy measured in diluted conditions for analogous systems. Unfortunately, given the complexity of the system, we cannot distinguish at this time between a single or double gap-filling phenomenon. Since each discontinuity is the site for a potential chemical reaction, we speculate here that a crowded supramolecular matrix produced by fragmented DNA double helices could provide a substrate where efficient primer extension and ligation reactions[21,22] can occur.

AUTHOR INFORMATION

Corresponding Author

Marco Todisco − *Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università di Milano, 20129 Milano, Italy;* Present Address: Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, Howard Hughes Medical Institute, 185 Cambridge Street, Boston, Massachusetts 02114, United States; orcid.org/0000-0001-6627-5283; Email: todisco@molbio.mgh.harvard.edu

Authors

Francesco Fontana − *Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università di Milano, 20129 Milano, Italy*

Tommaso Bellini − *Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università di Milano, 20129 Milano, Italy;* orcid.org/0000-0003-4898-4400

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.macromol.2c00856

Author Contributions

F.F. performed experiments, M.T. performed coarse-grained simulations, and all authors conceived the experiments and wrote the manuscript. All authors have given approval to the final version of the manuscript.

Funding

M.T. acknowledges the support from the Invernizzi Foundation. T.B. acknowledges support from MIUR-PRIN (Grant No. 2017Z55KCW).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) 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.

(2) Maffeo, C.; Luan, B.; Aksimentiev, A. End-to-end attraction of duplex DNA. *Nucleic Acids Res.* **2012**, *40*, 3812−3821.

(3) Kilchherr, F.; Wachauf, C.; Pelz, B.; Rief, M.; Zacharias, M.; Dietz, H. Single-molecule dissection of stacking forces in DNA. *Science* **2016**, *353*, No. aaf5508.

(4) 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.

(5) 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.

(6) 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.

(7) Hentschke, R.; Herzfeld, J. Isotropic, nematic, and columnar ordering in systems of persistent flexible hard rods. *Phys. Rev. A* **1991**, *44*, 1148−1155.

(8) Selinger, J. V.; Bruinsma, R. F. Hexagonal and nematic phases of chains. I. Correlation functions. *Phys. Rev. A* **1991**, *43*, 2910−2921.

(9) Nguyen, K. T.; Battisti, A.; Ancora, D.; Sciortino, F.; De Michele, C. Self-assembly of mesogenic bent-core DNA nanoduplexes. *Soft Matter* **2015**, *11*, 2934−2944.

(10) Zanchetta, G.; Bellini, T.; Nakata, M.; Clark, N. A. Physical polymerization and liquid crystallization of RNA oligomers. *J. Am. Chem. Soc.* **2008**, *130*, 12864−12865.

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

(12) Smith, G. P.; Fraccia, T. P.; Todisco, M.; Zanchetta, G.; Zhu, C.; Hayden, E.; Bellini, T.; Clark, N. A. Backbone-free duplex-stacked monomer nucleic acids exhibiting Watson−Crick selectivity. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E7658−E7664.

(13) 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. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1110−1115.

(14) Di Leo, S.; Marni, S.; Plata, C. A.; Fraccia, T. P.; Smith, G. P.; Maritan, A.; Suweis, S.; Bellini, T. Pairing statistics and melting of random DNA oligomers: Finding your partner in superdiverse environments. *PLoS Comput. Biol.* **2022**, *18*, No. e1010051.

(15) Zanchetta, G.; Nakata, M.; Buscaglia, M.; Clark, N. A.; Bellini, T. Liquid crystal ordering of DNA and RNA oligomers with partially overlapping sequences. *J. Phys.: Condens. Matter* **2008**, *20*, No. 494214.

(16) Todisco, M.; Smith, G. P.; Fraccia, T. P. Liquid Crystal ordering of DNA Dickerson Dodecamer duplexes with different 5′-Phosphate terminations. *Mol. Cryst. Liq. Cryst.* **2019**, *683*, 69−80.

(17) Walton, T.; Pazienza, L.; Szostak, J. W. Template-Directed Catalysis of a Multistep Reaction Pathway for Nonenzymatic RNA Primer Extension. *Biochemistry* **2019**, *58*, 755−762.
(18) Walton, T.; Zhang, W.; Li, L.; Tam, C. P.; Szostak, J. W. The Mechanism of Nonenzymatic Template Copying with Imidazole-Activated Nucleotides. *Angew. Chem., Int. Ed. Engl.* **2019**, *58*, 10812−10819.
(19) Zhou, L.; Kim, S. C.; Ho, K. H.; O'Flaherty, D. K.; Giurgiu, C.; Wright, T. H.; Szostak, J. W. Non-enzymatic primer extension with strand displacement. *Elife* **2019**, *8*, No. e51888.
(20) Ferris, J. P.; Hill, A. R., Jr.; Liu, R.; Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **1996**, *381*, 59−61.
(21) 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*, 7463.
(22) Todisco, M.; Fraccia, T. P.; Smith, G. P.; Corno, A.; Bethge, L.; Klussmann, S.; Paraboschi, E. M.; Asselta, R.; Colombo, D.; Zanchetta, G.; Clark, N. A.; Bellini, T. Nonenzymatic Polymerization into Long Linear RNA Templated by Liquid Crystal Self-Assembly. *ACS Nano* **2018**, *12*, 9750−9762.
(23) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011**, *134*, No. 085101.
(24) Sulc, P.; Romano, F.; Ouldridge, T. E.; Rovigatti, L.; Doye, J. P.; Louis, A. A. Sequence-dependent thermodynamics of a coarse-grained DNA model. *J. Chem. Phys.* **2012**, *137*, 135101.
(25) Snodin, B. E. K.; Randisi, F.; Mosayebi, M.; Šulc, P.; Schreck, J. S.; Romano, F.; Ouldridge, T. E.; Tsukanov, R.; Nir, E.; Louis, A. A.; Doye, J. P. K. Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J. Chem. Phys.* **2015**, *142*, 234901.
(26) Williams, S.; Lund, K.; Lin, C.; Wonka, P.; Lindsay, S.; Yan, H. In *Tiamat: A Three-Dimensional Editing Tool for Complex DNA Structures*, DNA Computing, Berlin, Heidelberg, 2009 , Goel, A., Simmel, F. C., Sosík, P., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; pp 90−101.
(27) Suma, A.; Poppleton, E.; Matthies, M.; Sulc, P.; Romano, F.; Louis, A. A.; Doye, J. P. K.; Micheletti, C.; Rovigatti, L. TacoxDNA: A user-friendly web server for simulations of complex DNA structures, from single strands to origami. *J. Comput. Chem.* **2019**, *40*, 2586−2595.
(28) Sengar, A.; Ouldridge, T. E.; Henrich, O.; Rovigatti, L.; Šulc, P. A Primer on the oxDNA Model of DNA: When to Use it, How to Simulate it and How to Interpret the Results. *Front. Mol. Biosci.* **2021**, *8*, No. 693710.
(29) Matt, J. Absolute Orientation - Horn's method (https://www.mathworks.com/matlabcentral/fileexchange/26186-absolute-orientation-horn-s-method), MATLAB Central File Exchange. **2022**, Retrieved April, 2022.
(30) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **2011**, *32*, 170−173.
(31) Roll, C.; Ketterle, C.; Faibis, V.; Fazakerley, G. V.; Boulard, Y. Conformations of nicked and gapped DNA structures by NMR and molecular dynamic simulations in water. *Biochemistry* **1998**, *37*, 4059−4070.
(32) Guo, H.; Tullius, T. D. Gapped DNA is anisotropically bent. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3743−3747.
(33) Sanstead, P. J.; Stevenson, P.; Tokmakoff, A. Sequence-Dependent Mechanism of DNA Oligonucleotide Dehybridization Resolved through Infrared Spectroscopy. *J. Am. Chem. Soc.* **2016**, *138*, 11792−11801.
(34) Chen, A.; Zhao, N. Comparative study of the crowding-induced collapse effect in hard-sphere, flexible polymer and rod-like polymer systems. *Phys. Chem. Chem. Phys.* **2019**, *21*, 12335−12345.
(35) 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.
(36) Di Leo, S.; Todisco, M.; Bellini, T.; Fraccia, T. P. Phase separations, liquid crystal ordering and molecular partitioning in mixtures of PEG and DNA oligomers. *Liq. Cryst.* **2018**, *45*, 2306−2318.
(37) Mariani, P.; Mazabard, C.; Garbesi, A.; Spada, G. P. A study of the structure of the lyomesophases formed by the dinucleoside phosphate d(GpG). An approach by x-ray diffraction and optical microscopy. *J. Am. Chem. Soc.* **1989**, *111*, 6369−6373.
(38) Mariani, P.; Spinozzi, F.; Federiconi, F.; Ortore, M. G.; Amenitsch, H.; Spindler, L.; Drevensek-Olenik, I. Guanosine Quadruplexes in Solution: A Small-Angle X-Ray Scattering Analysis of Temperature Effects on Self-Assembling of Deoxyguanosine Monophosphate. *J. Nucleic Acids* **2010**, *2010*, No. 472478.
(39) Izgu, E. C.; Fahrenbach, A. C.; Zhang, N.; Li, L.; Zhang, W.; Larsen, A. T.; Blain, J. C.; Szostak, J. W. Uncovering the thermodynamics of monomer binding for RNA replication. *J. Am. Chem. Soc.* **2015**, *137*, 6373−6382.

https://doi.org/10.1021/acs.macromol.2c00856
Macromolecules XXXX, XXX, XXX−XXX