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

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

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

Contributed by Noel A. Clark, May 31, 2018 (sent for review December 11, 2017; reviewed by Cyrus R. Safinya and Arjun G. Yodh)

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

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

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

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

Until now, stable duplex columnar ordering of NAs has required some degree of polymer linking of the stacking units, having been obtained as: (*i*) W-C duplexes (1) and columnar LCs (14) in NA polymers (Fig. 1 *D* and *I*); (*ii*) as W-C duplexes in mixtures of NA polymers and monomers (15–17); and (*iii*) as the chromonic columnar (COL) LC phase of ultrashort DNA and RNA oligomers (Fig. 1 *C* and *I*) (18–21). The latter are stabilized by the hierarchical self-assembly (HSA) of short complementary duplexes into linear aggregates via end-to-end attractive interactions (22–24), as a form of living polymerization (25–27), and the collective condensation of these aggregates into distinct birefringent LC droplet domains (18, 20, 22). Within these phase-separated LC domains, the long-ranged LC order creates a self-determined equilibrium state that performs molecular selection and can greatly enhance the effective concentration and stability of end-to-end oligomer contacts (18, 23, 25–27). As a result, the combination of selectivity, molecular organization, and fluidity uniquely provided by the LC ordering strongly promotes, with the addition of the appropriate chemistry, the covalent ligation of the oligomers of neighboring stacked duplexes into longer complementary strands (28). Such polymerization, in turn, generically stabilizes LC order, enhancing it in a positive feedback system of HSA, selection, and synthesis, which has been termed "liquid crystal autocatalysis" (LCA), and proposed as a path toward the prebiotic appearance of sequence-directed self-assembling linear polymers (28, 29). However, starting LCA from even few-base-long oligomers or by relying on polymer templates is problematical in a prebiotic context, since such molecular families are already highly selected and entropically narrowed, and could not have previously appeared without effective positive feedback and molecular selection (30). This motivates our search for (COL) chromonic LC order in the form of phase-separated LC domains capable of molecular selection and LCA in mixed starting populations of more primitive molecular species, such as NA monobases. Such lyotropic LC ordering has not been found to date in NA monobase variants without H bonding or with H-bonding

Significance

The columnar liquid crystal phases reported here are physical associations of the smallest molecular species to self-assemble into the duplex base-paired stacked columnar double-helical structures of nucleic acids. These assemblies of monomers can provide starting states capable of partitioning appropriate molecules from solution with a high degree of selectivity, acting as pathways for the prebiotic appearance of molecular selection, self-assembly, and, ultimately, of the sequence-directed assembly of polymers.

Author contributions: G.P.S., T.P.F., M.T., G.Z., T.B., and N.A.C. designed research; G.P.S., T.P.F., M.T., G.Z., C.Z., E.H., T.B., and N.A.C. performed research; G.P.S., T.P.F., M.T., G.Z., C.Z., E.H., T.B., and N.A.C. analyzed data; and G.P.S., T.P.F., T.B., and N.A.C. wrote the paper.

Reviewers: C.R.S., University of California, Santa Barbara; and A.G.Y., University of Pennsylvania.

The authors declare no conflict of interest.

[1]G.P.S. and T.P.F. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: tommaso.bellini@unimi.it or noel.clark@colorado.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1721369115/-/DCSupplemental.

Published online July 2, 2018.

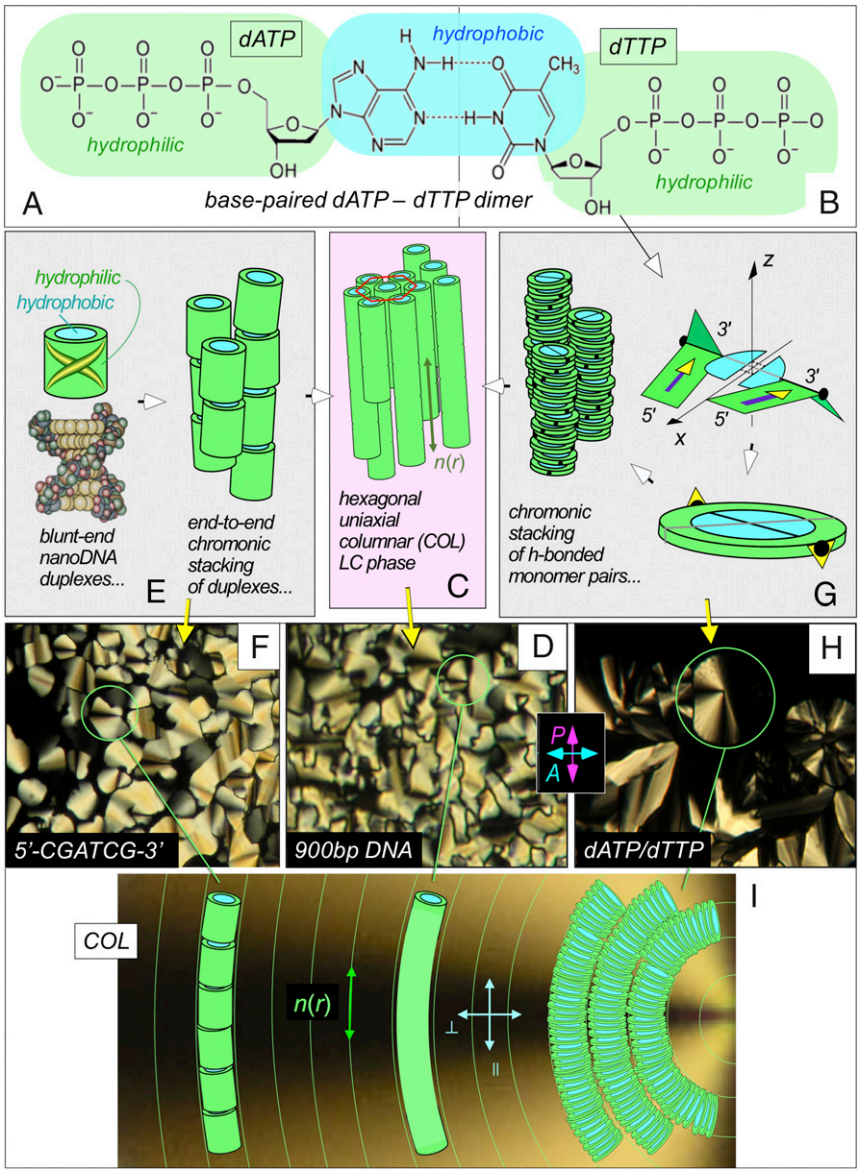

Fig. 1. Lyotropic chromonic LC ordering of nucleoside triphosphates. (*A* and *B*) Chemical structures of dATP and dTTP and of a W-C hydrogen-bonded dATP/dTTP dimer showing its hydrophobic and hydrophilic domains. The hydrogen bonding reduces the solubility of the base pair, promoting the chromonic stacking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

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

Results

Liquid Crystal Ordering of NTPs in Aqueous Solution. We report here the duplex [COL chromonic HSA and LC ordering of the mononucleoside triphosphates, chosen over other mononucleoside species because of their enhanced aqueous solubility and potential for ligation provided by the triphosphate tail (40)]. The molecules investigated were the DNA mononucleoside triphosphates dATP, dCTP, dGTP, dTTP, and dUTP, and the RNA mononucleoside triphosphates rATP, rCTP, rGTP, rTTP, and rUTP. LC phase behavior in aqueous solution and the structure of the LC COL ordering was systematically studied by polarized transmission optical microscopy (PTOM) and X-ray diffraction (XRD) vs. temperature T, and NTP anion concentration, c(mg/mL) (*SI Appendix*).

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

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

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

Several rNTP and rNTP/dNTP binary mixtures were also investigated at temperatures $T >$ −20 °C, with the results summarized in the matrix of observations of the COL phase shown in *SI Appendix*, Fig. S16. Of note is the appearance of the COL phase in rCTP/rGTP and rATP/rTTP mixtures, but not rATP/rUTP mixtures. The latter absence is an observed break from the

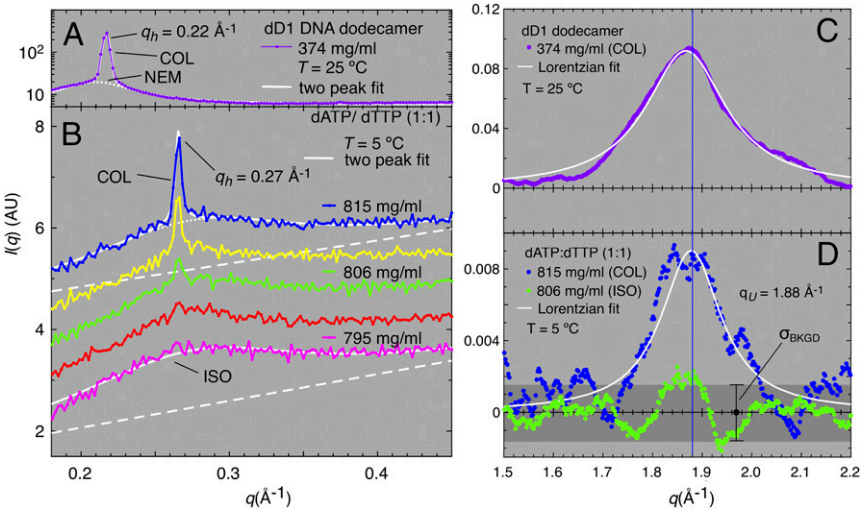

Fig. 2. X-ray structure factors *I*(*q*) showing the periodicities in the COL LC phase due to the packing of duplex columns of chromonic aggregates and base stacking within the columns. (*A*) The dD1 dodecamer, included for comparison purposes, forms chromonic COL stacks of blunt-end duplexes, which, at *c* = 374 mg/mL order into coexisting uniaxial hexagonal COL and uniaxial nematic (NEM) phases. (*B*) Coexisting COL and ISO phases versus concentration *c* in the dATP/dTTP (1:1) mixture. These data are taken along the black box in *SI Appendix*, Fig. S9*C*. The diffuse peak in the ISO phase indicates short-ranged aggregation and packing of columns. The ISO peak is similar to that found in chromonic phases of organic dyes (65). (*C* and *D*) X-ray structure factors *I*(*q*) at large *q*, showing the base-stacking periodicity in the chromonic aggregates forming COL LC phases. The base-stacking peak is at $q_U = 1.88$ Å^{-1} in both dD1 and the NTP mixtures, corresponding to $L_U = 3.34$ Å. (*B* and *D*) Both the column ordering and base-stacking peaks disappear upon entering the ISO phase (green). (*D*) σ_{BKGD} indicates the shot noise of the background level.

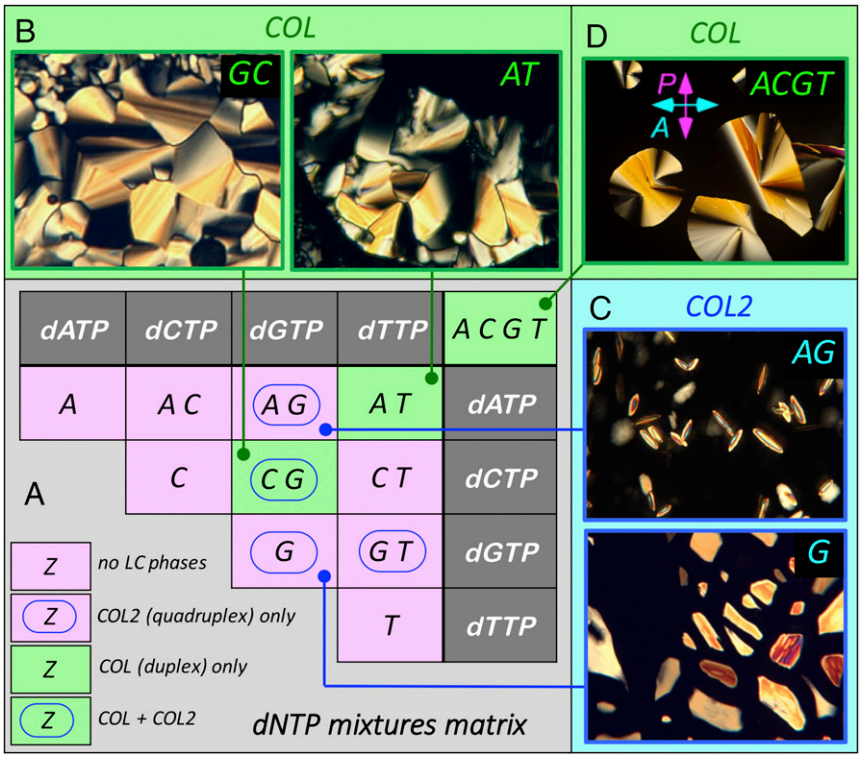

Fig. 3. ACGT matrix summarizing the search for LC phases in aqueous solutions of dNTPs and their mixtures, showing a significant dependence on choice of base. (*A*) The matrix colors indicate the most ordered phase to appear for each base-pair combination for $T > -20$ °C. The COL phase appears in NTP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

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

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

Using this assumption and the melting entropy (46–48) gives a reasonable estimate of this depression (*SI Appendix*), indicating the selectivity of the phase separation between a stoichiometrically balanced 1:1 COL LC domains and an A-rich isotropic phase.

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

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

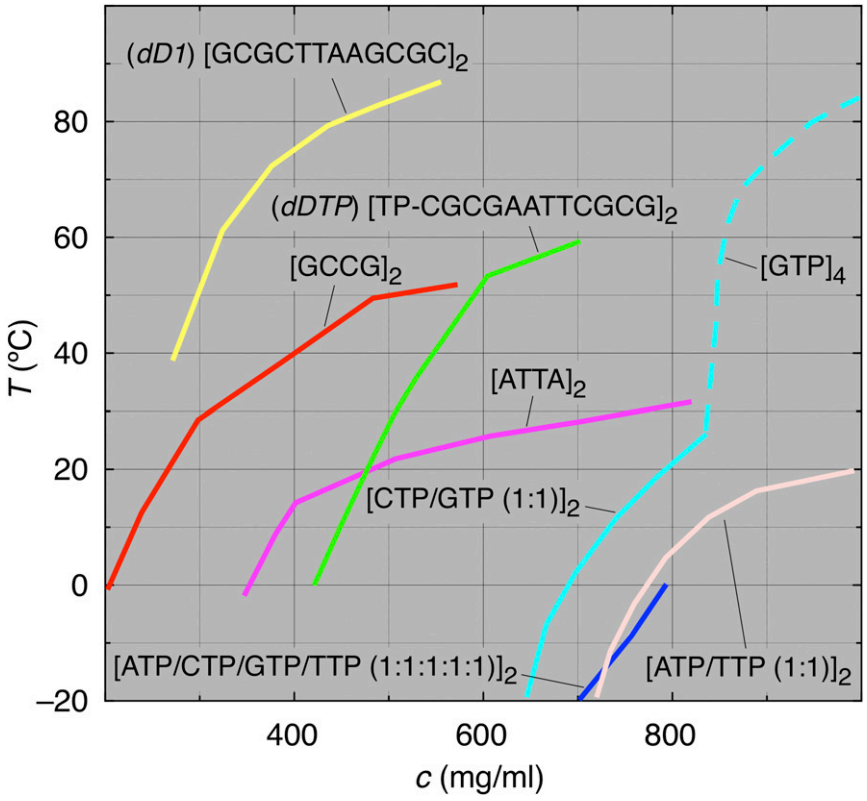

Fig. 4. Compilation of relevant oligomer and monomer DNA LC phase diagrams, determined by PTOM and XRD (*SI Appendix*, Figs. S7 and S9 and *SI Appendix*), showing the high-temperature extent of LC ordering in: blunt-end self-complementary dodecamer duplexes (dD1); triphosphate-terminated self-complementary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

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

A schematic of a proposed W-C arrangement of the ATP/TTP COL phase is shown in Fig. 1*G*. The arrangement of the hydrogen-bonded bases and the antiparallel stacking of the 5′-3′ direction of the ribose rings is similar to that of polymer B DNA. In B DNA the base stacking is helixed because of the bulkiness of the ribose and the phosphodiester bond, and the convenient length of the latter. While the phosphodiester bond is absent in the NTP duplex stacks, the triphosphate group, having a molecular volume (~400 Å^3 per base) that cannot be accommodated in any simple vertical stacking, drives, along with the ribose, the tendency for adjacent base pairs to a nonzero relative azimuthal orientation about the helix axis, z (Fig. 1*G*). The ribose chirality biases this relative orientation of bases: (*i*) toward helixing; and (*ii*) toward the antiparallel orientation of the two 3′-5′ ribose directions at a given level, making them symmetrically related by a π-rotation about $\boldsymbol{x}$ and therefore of the same handedness (Fig. 1*G*), which, as in polymer DNA, is required if the molecular conformations in the helical windings of the two phosphodiester/ribose chains are to be of the same average configurational structure. We thus propose a local structural organization very similar to that of the W-C of dilute polymeric duplex DNA, the double helix with triphosphates replacing the phosphodiester bonds of the polymer, as drawn schematically in Fig. 1*G*. In any case, whatever the details of the base-pairing geometry, the W-C selectivity indicates that that the COL LC ordering is a sensitive selection tool, capable on its own of stabilizing certain base-paired duplex DNA structures and destabilizing others in both the concentrated COL LC phase of monomers and the dilute duplex polymer chains.

Comparative Phase Behavior of NA Oligomers and Monomer Nucleotides. The optical and X-ray data enable the construction of LC phase diagrams for the two stoichiometrically balanced dATP/dTTP (1:1) and dCTP/dGTP (1:1) mixtures, shown in Fig. 4 and *SI Appendix*, Fig. S9, and their comparison with those of the four-component mixture, dATP/dCTP/dGTP/dTTP (1:1:1:1) and of a variety of previously studied complementary nanoDNA oligomers: DNA dodecamer 5′GCGCTTAAGCGC3′ (dD1); 5′ triphosphorylated DNA dodecamer TP-5′CGCGAATT CGCG3′ (TPdD1); and the DNA 4mers dATTA and dGCCG

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

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

Discussion

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

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

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

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

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

Materials and Methods

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

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

ACKNOWLEDGMENTS. This research was supported in part by the NSF Biomaterials Program under Grant DMR-1611272, by NSF Materials Research Science and Engineering Center Grant DMR-1420736 to the University of Colorado Soft Materials Research Center, by NIH/University of Colorado, Boulder Molecular Biophysics Training Program support of G.P.S., and by a grant to T.P.F. from the John Templeton Foundation, provided through the Earth-Life Science Institute of the Tokyo Institute of Technology.

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