Supporting Information

Bellini et al. 10.1073/pnas.1117463109

SI Materials and Methods.

Random DNA (ranDNA) Synthesis. The largest part of random oligoDNA was synthesized using an Äkta Oligopilot, manufactured by Amersham Pharmacia Biotech. The majority of coupling reagents were ordered from Sigma-Aldrich. Amidites, solid support (nucleotide bound on polystyrene), and ethylthio tetrazole were purchased from American International Chemical, Inc. Randomness of the synthesized oligomers at the designed positions in the sequences was obtained by providing the oligopilot with a single vessel containing an equimolar solution of each of the four phosphoramidites. Phosphoramidites were purchased and utilized only if both coupling efficiencies and purity level were greater than 99%, to ensure no reactivity differences. Given equivalently high levels of both purity and reactivity for each individual phosphoramidite, it is safe to conclude that no significant bias toward any specific base is expected in the syntheses of oligomers when the oligopilot is provided an equimolar solution of all of the phosphoramidites.

In the last step of synthesis, we omitted removal of the 4,4′-dimethoxytrityl (DMT) blocking group to facilitate later purification. Materials synthesized using the Äkta Oligopilot came off the synthesizer as bound oligoDNA on solid support. OligoDNA was cleaved from the solid support by stirring the solid support-bound oligoDNA in an ammonium hydroxide solution (ammonia content approximately 28–30%) in a ratio of 40 mL per 1.6 g of solid support material. The suspension was heated to 50 °C for 8 h and then allowed to cool to room temperature. The cooled suspension was vacuum filtered over a coarse frit and the solid filtrate washed with three portions of an ethanol/water solution (1 : 1 by volume, total volume of 400 mL). The resulting solution was then stripped of solvent under vacuum to provide the crude unpurified oligoDNA.

Crude oligoDNA was purified using preparatory HPLC. A dual Varian Prostar 218 system tied to a Prostar 318 detector employing a SepaxGP-C8 column from Sepax Technologies Inc. was used to remove failure sequences from the crude oligoDNA to yield purified random 20-meric oligoDNA. The oligoDNA was purified by the "DMT-on" HPLC purification method. Crude oligoDNA was suspended in 50 mM triethylamine bicarbonate (TEAB) buffer with 100 mM sodium chloride and filtered through a 0.2-μm nylon syringe filter to remove particulates. OligoDNA was then passed through the HPLC using 50 mM TEAB as the primary buffer and HPLC grade acetonitrile as the secondary buffer. DMT-capped oligoDNA separates from failure sequences on the column using a slow linear elution gradient with increasing acetonitrile concentration. The presence of the DMT cap causes the sequence of interest to elute much later than the uncapped contaminating sequences, giving efficient separation. Once eluted from the HPLC column, the polynucleotide was concentrated to dryness using a rotary evaporator. Then 40 mL of 80% glacial acetic acid with 20% water was added to the dried oligoDNA and mixed one-half hour at room temperature, with shaking, to cleave the DMT group from the 5′ end of the polynucleotide. Acetic acid was predominantly removed by evaporation on the rotary evaporator and then the oligoDNA was resuspended in 40 mL of 50 mM TEAB with 2% ammonium hydroxide by sonicating at room temperature for several minutes. A syringe filter was again applied to remove particulates so that the solution could be sent through the HPLC a second time using the same column, buffers, and elution gradient as the initial run. In the second run, the deprotected oligoDNA elutes from the column at the same hydrophobicity as the failure sequences eluted in the initial run. Following elution from the column, the oligoDNA was concentrated to dryness using the rotary evaporator. At this step, the purified DNA can be dissolved in water and used as necessary for other experimentation. MALDI-TOF mass spectrometry, as described elsewhere, was used after each HPLC step to confirm the presence of the expected oligoDNA.

Parts of the sequences—HPLC purified—were purchased from PRIMM as liophylized powders.

ranDNA Characterization by HPLC and MALDI Mass Spectroscopy. Initial characterization was performed by HPLC. We show in Fig. S1 the HPLC traces obtained with a single distinct 12-mer DNA (CGCGAATTCGCG, Dickerson Dodecamer, 12SC) and of the fully random 20-mer (20*N*). As compared to the purification of 12SC, HPLC traces of the final step of 20*N* purification have a broader elution profile which displays characteristics of a composite peak. This behavior would be expected of a sample containing many different DNA sequences of a characteristic length but each eluting at a slightly different condition of hydrophobicity due to the variations in A, T, C, and G content. It also eluted somewhat later than the 12SC, suggesting a uniformity of elution depending on the number of bases present in all of the 20*N* sequences.

The 20*N* DNA was further characterized by MALDI-TOF mass spectrometry as shown in Fig. S2 (left axis). In order to estimate the width of the expected mass distribution for 20*N*, we generated 10^8 random 20-mer sequences on the computer using the Random Number Generator by Wolfram Mathematica and determined their expected mass distribution $P(M)$, shown in Fig. S2 (right axis). The overplot on the figure enables comparison between the measured and the calculated mass distribution. The MALDI peak was approximately $6{,}110 M_r$ as compared to the calculated theoretical average weight of $6{,}160 M_r$. The added broadness at the lower molecular weight fringe of the peak is likely more populated than the theoretical average as a result of some small-scale damage to the polymer done in the process of laser ablation. The purified 20*N* showed an additional contaminant peak at slightly higher molecular weight, suggesting the inclusion of some 20*N* still containing the 5′-terminal protecting group left appended for purification.

Sample Preparation. Semiconcentrated (10–100 mg/mL) solutions were prepared in milliQ water and loaded between parallel plate glass plates with gaps in the range 5–20 μm. Higher concentrations were obtained by evaporation; the solution drop was then sealed with fluorinated oil, which warranted cell stability for some days. For centrifugation experiments, DNA solutions were loaded into glass cylindrical capillaries (0.53 mm inner diameter) and flame-sealed.

Microscopic observations were made either on a Nikon TE200 or on a Nikon Optiphot2 microscope and images were acquired on Nikon DS-5M and Prosilica GX-1050 b/w CCD cameras.

Thermal history effect and determination of the liquid crystal (LC) volume fraction. All the samples were subjected to a preliminary annealing step to homogenize concentration: the cell was kept for 1 h at 90 °C and then slowly cooled down to room temperature in approximately 4.5 h (red line in Fig. S3).

To explore for thermal history dependence, we carried out a set of measurements after and during thermal cycles between three significant temperatures:

1. 25 °C, at which both helical structures and LC phases are stable;
2. 40 °C, where the LC phases are completely melted at all the investigated concentrations, whereas most (≥80%) of the 20*N* duplexes in solution are bound;
3. 90 °C, where all the DNA helices are melted and the system is a homogeneous solution of single-stranded filaments.

In a typical cycle the sample was heated from room temperature up to 90 °C and then cooled at fixed temperature rates in two steps, 90 → 40 °C—where the formation of helices takes place—and 40 → 25 °C—where most of the helices are formed and nucleation occurs. To explore the possible effects of thermal history on the quality of helices pairing and on the LC phases features, we varied the cooling rates in the range 0.01–50 °C/min, corresponding to residence times in the 90 → 40 °C and 40 → 25 °C intervals of 1–5,000 and 0.3–1,500 min, respectively.

The evolution of LC textures was monitored through depolarized transmitted light microscopy (DTLM), with LC domains emerging as bright spots from the dark, isotropic background (see Fig. 2*B* in the main text). Thresholded 8-bit images were analyzed with ImageJ to measure the total volume fraction of the LC phase and the number and the average size of the domains; extracted quantities were plotted vs. time at fixed T or vs. temperature at fixed rate. As expected from a nucleation and growth process, after a lag time the LC volume fraction grows from 0 to a stable ϕ_{LC} (Fig. S4), whereas the number of domains reaches a maximum in a few minutes and then decreases because of merging.

From careful characterization at different DNA concentrations and cooling rates, we found that

i. for a given concentration, ϕ_{LC} measured after long equilibration at 25 °C is basically independent from both the 90 → 40 °C and the 40 → 25 °C cooling rates, the variations in the resulting ϕ_{LC} contributing to the vertical error bars in Fig. 3*A*;
ii. LC domain number and size depend on the cooling time from 40 to 25 °C, with faster cooling (>0.5 °C/min) resulting in the nucleation of a larger number of smaller LC domains;
iii. columnar (COL) domains coexisting with isotropic, after cycles 25-40-25 °C, form in roughly the same locations, whereas cycling 25-90-25 °C makes the system lose memory of the location of COL domains;
iv. samples held at 25 °C for a longer time have a better thermal stability (see Fig. 3*B*).

Concentration Measurements via Microscope-Based interferometry. The measurement of DNA concentration in sealed cells was performed by microscope-based interferometry. Planar cells made with high-refractive index glass (F2 Schott, $n = 1.62$) were brought, through heating, into the isotropic phase and the spectrum of the light reflected by the illuminated portion of the cell was collected. The multiple reflection fringes due to the interference from the parallel glass plates were systematically measured in various spots within the DNA solution and on the periphery of the cells, where the fluorinated oil was instead present. This set of data enabled determining the refractive index of the DNA solution and thus the DNA concentration c through a calibration curve ($n = n_0 + dn/dc \cdot c$, with n_0 the refractive index of the solution at zero DNA concentration and $dn/dc = 0.136$ cm^3/g). The extracted values have an uncertainty of about ±50 mg/mL.

Capillary Centrifugation and Concentration Measurements via UV Absorption. A calibration curve was created for determining random 20-mer concentration from the absorbance at 260 nm measured by spectrophotometry. Small quantities of between 1 and 3 mg of solid DNA were weighed on a Toledo-Mettler scale and then diluted in 0.5 mL of pure water. A dilution series was produced by adding water at 10% of the current volume of the sample to the sample at every measurement step. At each dilution, the sample's absorbance was measured with a Hewlett Packard 8452A Diode Array Spectrophotometer in a 0.7-mL quartz cuvette with a 1-cm optical path length. The method of first diluting the sample and then measuring the absorbance was repeated until the absorbance of the sample dropped below measurable values. The entire series was repeated several times with available DNA to minimize the numerical error associated with the weight measurement on the scale and with the spectrophotometry.

Sealed capillary containing DNA liquid crystal samples were centrifuged in a Sorvall RC70 centrifuge with an AH-629 rotor. Samples typically were treated with a 10-min 40 °C melt followed by a slow overnight cooling step back to room temperature to anneal. Samples were centrifuged at the ambient temperature of storage upon completion of the annealing step, typically 21–22 °C, because variations in temperature during centrifugation by as little as a degree or two produced additional liquid crystal in the capillary. The centrifuge generated sufficient force to sediment liquid crystal in the sample at 10,000–12,000 rpm for as little as 30 min, though additional 30-min steps were applied to sediment straggling domains. Domains adhered to the glass walls of the capillary typically came free at lower than 15,000 rpm, though individual domains were sometimes seen floating at the meniscus. Relatively gentle centrifugal forces were employed in an effort to avoid disrupting existent domains in order to ensure that sedimentation was dependent on domain density and not on molecular sedimentation coefficients.

After centrifugation, capillaries were sectioned carefully by a glass cutter, cutting first at the meniscus and then once or twice in the isotropic region and finally once at the isotropic/liquid crystal interface near the bottom of the capillary. The cutting must be performed in that order to prevent the liquid from migrating toward the air bubble in the capillary during cutting, thus disturbing the correlation between liquid length and glass length needed for later volume measurement. The glass sections containing the sample were placed in 1.7-mL Eppendorf tubes where they were immersed in 1 mL of pure water. In early trials, samples in the glass sections were displaced and dissolved by a quick sonication step, but this step was later determined to be unnecessary. Aqueous samples were further diluted in water until the 260-nm absorbance was at about a 0.2-AU level as measured with the Hewlett Packard 8452A spectrophotometer.

By relying on the previously determined calibration curve, we could correlate absorbance to the dissolved DNA quantity and thus determine the average concentration of DNA in each glass section. Glass sections were dried and then measured with Mitutoyo precision calipers in order to determine the volumes of each. Concentration was recorded from the absorbance calibration, and the LC fraction was recorded by direct measurement of the volume of the LC as a fraction of the volume of the total sample, including LC and isotropic fractions both.

Melting Temperature Measurements. The duplex denaturation transition was probed by measuring the fluorescent emission of ethidium bromide (EtBr) mixed into DNA solutions at a concentration of one molecule per duplex. Upon slowly increasing T, as the duplexes unbind, the EtBr fluorescence is greatly reduced and eventually reaches a high-T plateau, enabling the determination of the fraction of paired strands. The melting temperature is defined as that corresponding to half of the population of unpaired helices.

Measurements of the Melting Temperature of the Columnar Phase. The LC domains melt as T is raised. The melting of the columnar LC phase has been characterized by recording DTLM images

as temperature is raised at a constant slow rate (0.2–0.5 °C/ min). As described above, the total volume fraction of the LC phase (ϕ_{LC}) is determined via software analysis of thresholded 8-bit images. ϕ_{LC} is plotted vs. time or vs. temperature, and columnar melting temperature (T_{COL}) is defined as the temperature for which $\phi_{LC} = 0$. $\phi_{LC}(T)$ melting curves are found to depend on the thermal histories of the samples. LC domains that previously had a short equilibration time or a fast cooling rate melt in a shorter time showing steep melting curves. On the other hand, LC domains that were kept for longer time at $T = 25$ °C melt in a longer time and at higher temperature (compare curves 2 and 3 in Fig. 3*B* of the main text). The measured T_{COL} for 20*N* are spread in the interval between 28 and 36 °C, while self-complementary sequences are characterized by a higher and well-defined T_{COL} ($T_{COL} \approx 49$ °C for 12SC and $T_{COL} \approx$ 78 °C for 20SC).

Study of the Effects of Partial Complementarity on LC Ordering.

Table S1 shows that even a single terminal mispairing impairs the self-association of the duplexes, as might be expected. Sequences paired "in-register" that have one or two noncomplementary bases in the one terminal do not form nematics, whereas defects in both terminals suppress both LC phases. Internal errors are less problematic, affecting the phase behavior only when there are more than two in a duplex, in which case the nematic phase is suppressed and the COL phase is still found. In solutions of sequences paired in a shifted mode with overhangs, LC ordering is stabilized if the overhangs are mutually complementary but suppressed if the overhangs are noncomplementary. Ionic strength of Na^+ does not affect significantly this behavior, in line with previous observations.

Phase Behavior of Mixtures of Interacting/Noninteracting Duplexes.

In order to test the generality of the polydispersity of interduplex interaction as a mechanism determining the phase behavior, we studied mixtures of helices capable of linear aggregation with helices with noninteracting terminals (sketched in Fig. S5*A*). These were obtained with two different sets of sequences:

- 12SC-CG: 5′-CGCGAATTCGCGCG-3′, which forms helices with self-complementary unpaired ends, favoring end-to-end aggregation;
  A: 5′-CCTCAAAACTCC-3′;
  FITC-B: FITC-5′-GGAGTTTTGAGG-3′, complementary to A and with a bulky fluorescent moiety at 5′ terminal which prevents linear aggregation and LC formation of the resulting helices);
- 12SC′-CG: 5′-GCTAGCGCTAGCCG-3′, forming helices with sticky ends;
  12SC-TT: 5′-CGCGAATTCGCGTT-3′, which instead yields noncomplementary dangling ends and does not show LC behavior.

In both mixtures coexistence between birefringent COL domains and isotropic solution was observed. For set A, the FITC group, while disturbing the linear aggregation, also allows to detect the location of A/FITC-B duplexes through its fluorescence emission. Fig. S5 *C* and *D* show a DTLM picture of such a mixture and the related fluorescence microscopy image: the birefringent LC domains correspond to low fluorescence regions, richer in the interacting, non fluorescent 12SC-CG, whereas the isotropic background has a higher fluorescence, signaling a higher concentration of labeled, noninteracting duplexes.

As for set B, we mixed the two duplex families by varying their relative ratio and measured the resulting LC fraction, shown in Fig. S5*B*. There is a clear correlation between the fraction of interacting duplexes in solution and the average LC occupancy, demonstrating that the LC domains are mainly populated by interacting duplexes.

Estimate of the Interduplex Interaction Free Energy from the Phase Diagram.

To estimate the interaction free energy between neighboring helices from the measured phase diagram of ranDNA, the following procedure was used. From the theoretical and simulated hard rods phase diagram we know that the predicted critical volume fraction for the transition to nematic is $\phi_{IN} \sim 4D/L$ (1, 2), where L is the length and D the diameter of the rods (the linear aggregates in our case). Therefore, from the lowest ϕ_{IN} measured at 25 °C for ranDNA we can estimate the aspect ratio of the aggregates at the transition and thus their aggregation number M. Various models correlate M with the interaction energy between duplexes. Following refs. 3 and 4, the aggregation number of duplexes in an aggregate depends on $\varepsilon \equiv \Delta G/k_BT$ and on ϕ as

$$M = \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\varphi e^{\varepsilon + k_I \varphi}}$$

where $k_I \sim 1.45$ is a virial coefficient taking into account the steric repulsion between monomers. Furthermore, for a given ϕ, at increasing T we can assume that the LC phase is lost when the length of the aggregates falls below the Onsager limit and thus we can extract ΔG from the measured melting temperature of the LC domains. The results of such analysis are reported in Table S2: considering the uncertainty in the LC melting temperature due to thermal history dependence and the uncertainty in the concentration, we obtain from the set of data points reported in Fig. 3 *A* and *B* an average value of ΔG between 1.4 and 2.2 kcal/mol (approximately 2.3–3.7k_BT). This value is significantly smaller than the one obtained for both blunt ends ($5k_BT$) and sticky ends of length $\ell = 2$ ($6k_BT$), as expected on the basis of the melting behavior of Fig. 3*B* of the main text. However, this value is about double the value obtained from the computer estimate. This discrepancy has to be ascribed to the fact that the free energy values used to compute the thermodynamic stability of DNA double helices have a large degree of uncertainty (5). Moreover, such values often are found to be smaller when compared to those determined through the analysis of DNA LC phase diagrams, as can be appreciated by comparing the stacking energy determined from the experimental behavior of LC made of blunt end helices ($5k_BT$) and the average free energy per quadruplet from ref. 6 ($2.7k_BT$), which should in principle be similarly determined from base stacking.

Computer-Based Estimates of the Statistical Properties of ranDNA.

Adopted Assumption. DNA double helices are formed by two distinct DNA strands held together by the combination of pairing and base stacking forces. These forces control the selectivity of the binding process, by which the binding energy depends on the degree of complementarity of the nucleobase sequences along the two polymers. Base pairing originates from the formation of hydrogen (H) bonds between pairs of nucleotides. The strength of the interaction is maximum when the sequences match according to the Watson-Crick (WC) complementary couples—i.e., adenine (A) with thymine (T) (two H bonds) and guanine (G) with cytosine (C) (three H bonds). The energy associated with the formation of the double helix ("hybridization") depends on the length of the nucleic acid polymer and on the quality of the WC matching, vanishing for pairs of sequences with poor complementarity. Base stacking forces, instead, are attractive forces between the aromatic hydrocarbon plates of the nucleobases. Stacking is much less selective than pairing and its strength mainly depends on the overlap between the aromatic plates.

The most commonly adopted approach to evaluate and predict the overall binding free energy ΔG between hybridized DNA strands relies on the so-called "nearest neighbor" (NN) model. According to it, the double helix is decomposed in quadruplets formed by two consecutive nucleotides on one strand and the corresponding nucleotides on the other strand, as it is shown in Fig. S6*A*, where two partially complementary 10-mers are taken as an example. Ten quadruplets can be distinguished, that include also the two "triplets at the extremes". Notice that the quadruplets overlap.

The NN model neglects contribution arising from nonlocal (beyond nearest neighbors) interactions along the helix. According to this model, the total free energy involved in the formation of the double helix can be estimated by $\Delta G = \sum \Delta G_Q + \Delta G_{\mathrm{INIT}}$, where the summation is on all the quadruplets of the sequence, ΔG_Q is the free energy of each quadruplet, and ΔG_{INIT} is an "initiation" free energy. The analysis of the database in ref. 6 indicates that ΔG_Q significantly promotes the formation of the double helix only when the quadruplet is formed by two WC pairs. When averaged over all the possible quadruplets involving WC pairs (AA/TT, AT/TA, TA/AT, CA/GT, GT/CA, CT/GA, GA/CT, CG/GC, GC/CG, and GG/CC), $\langle \Delta G_Q \rangle \sim -1.6$ kcal/mol at $T = 25$ °C, corresponding to approximately $2.7 k_{\mathrm{B}}T$. Pairing errors in one of the two couples lead, on average, to an increased free energy: $\langle \Delta G_Q \rangle \sim 0.4$ kcal/mol.

The computer-based free energy evaluation reported in this study is based on the NN model which is further simplified by the following assumptions:

i. We neglect differences in free energy contributions among well paired quadruplets. Accordingly, quadruplets 3, 4, 5, 8, and 9 all give the same contribution $\Delta G_0 = \langle \Delta G_Q \rangle$ to the binding free energy, marked, in Fig. S6*B* and in Fig. 4*C* of the main text, with staple shapes below the duplex.
ii. We neglect the contribution to the free energy of quadruplets in which one or both pairs are not WC pairs. Among these are the Hoogsteen bonds. Our neglecting these contributions is based on the necessity of defining a simplified approach and on the smaller energies involved in non-WC pairs, as evident from the analysis of ref. 6. Thus, the contribution of quadruplets 2, 6, and 7 to the duplex binding free energy is set to 0.
iii. We neglect the contribution to the duplex stability of "coaxial stacking", the situation occurring when one base is overhanging at the end of a duplex (triplets 1 and 10), always quite smaller than ΔG_0.
iv. We neglect ΔG_{INIT}. Among the approximations here introduced, this one is the most significant. A better approximation would have been to set $\Delta G_{\mathrm{INIT}} = \Delta G_0$, because their values are comparable. Introducing this amelioration in the calculations brings to a shift of $-1\Delta G_0$ in the free energy scales of Fig. 4 in the main text and a different value for the attempt rate in the duplex lifetime equation—i.e. $\tau_0 \sim 150$ ns. However, we decided not to include this variant because it makes the discussion more complex without introducing any real change in the results. Furthermore, ΔG_{INIT} has been well investigated for duplexes with a large fraction of WC pairing, and it may not simply translate to situations as those explored here, where the degrees of freedom within the duplexes are much larger because of the high degree of pairing errors.

Evaluation of Intraduplex Binding Free Energy. These simplifying assumptions enable a prompt estimate of the duplex pairing free energy, as in the examples in Fig. S6*C*, where we consider the same motifs of partial pairing introduced in Fig. 1 of the main text. This simplified approach to the computation of duplex binding free energy has enables us to perform statistical evaluation on large numbers of computer generated partially paired duplexes, as described in the following.

Evaluation of the Interduplex Interaction Free Energy for Dickerson Dodecamer with Random Overhangs. For each length R of random overhangs, 10^7 collisions of randomly generated overhangs were considered. For each, the value of free energy has been determined as described above. This process is sketched in Fig. S6*D*. For $R > 2$ shifted configurations were also included (as in the last example of Fig. S6*D*. The resulting values for $\langle \Delta G \rangle$ are reported in the main text and in Fig. 4*D* (right axis).

Evaluation of $n(\Delta G)$. Within the ensemble of fully random sequences of a given length, $n(\Delta G)$ expresses the number of possible oligomer combination yielding a given duplex binding free energy ΔG. As described in the main text, random collisions between randomly generated sequences yield a large variety of pairing motifs, some of which are sketched in Fig. S6*C* for fully random 10-mers (10*N*). An important point is that by random collision we mean that we evaluate the energy not only for in-register couples of oligomers (potentially leading to blunt-ended duplexes if terminals produce WC pairing) but also for oligomers that are brought in contact with a shift of arbitrary (randomly determined) length, a situation giving rise to duplexes with overhangs.

More specifically, for the case of 20*N*, $n(\Delta G)$ as reported in Fig. 4*A* of the main text is expressed as the number of combination that one given 20-mer sequences can make with the other ones of the 20*N* family so to have a binding strength equal to ΔG. Accordingly, $n(19\Delta G_0) = 1$, because in the 20*N* pool one only sequence is the perfect complementary of the starting sequence. Analogous definitions apply to the 6*N* and 12*N* cases reported in the figures.

$n(\Delta G)$ has been generated by merging its behavior at small and large ΔG that were independently evaluated. By generating, at the computer, a number N_{SIM} of duplexes in the range 10^7–10^8 and evaluating the binding free energy for each of them, we have obtained the free energy distribution at low ΔG. This approach is necessarily limited to small ΔG because the number of sequences that can be generated is small compared to the $N_{\mathrm{TOT}} = 4^{20} \approx 10^{12}$ sequences of the 20*N* system. In this small ΔG limit, $n(\Delta G)$ is obtained by multiplying the number of computer generated duplexes at each value of ΔG by the factor $N_{\mathrm{TOT}}/N_{\mathrm{SIM}}$—i.e., the ratio between the number of different sequences actually present in a ranDNA system and the number of sequences generated by the computer.

The behavior of $n(\Delta G)$ for large values of ΔG can instead be analytically determined. For example, in the case of 20*N*, by examining the possible source and location of errors in pairing, it can be recognized that $n(19\Delta G_0) = 1$, $n(18\Delta G_0) = 14$, $n(17\Delta G_0) = 161$, $n(16\Delta G_0) = 1{,}681$, and $n(15\Delta G_0) = 10{,}970$. These values are the rightmost blue squares in Fig. 4*A*. As is visible in the figure, despite the gap between the two approaches [$n(13\Delta G_0)$ and $n(14\Delta G_0)$ could not be evaluated either analytically or with computer simulation], the two different determinations of $N(\Delta G)$ very well match and are perfectly fitted by the product of an exponential decay and a Gaussian.

The same combination of approaches enabled us to determine $n(\Delta G)$ for the 6*N* and the 12*N* ranDNA solutions reported in Fig. 4*A*.

Evaluation of the Equilibrium and Nonequilibrium Distribution in Duplex Binding Free Energy $P(\Delta G)$. Equilibrium distributions are obtained by multiplying $n(\Delta G)$ and the Boltzmann factor $\exp(\Delta G/k_{\mathrm{B}}T)$, both plotted in Fig. 4*A*. The resulting distributions are shown, after normalization, in Fig. 4*B*.

Kinetic distributions are determined via computer simulation by following the time evolution one given sequence through the

consecutive formation of duplexes with other randomly chosen strands. Specifically, the nonequilibrium distribution is obtained through the following steps:

i. one specific ("original") sequence is randomly chosen and held;
ii. randomly chosen partner sequences are produced, and their binding energy with the original sequence are determined, as discussed above, for given randomly chosen shifts;
iii. on the basis of the binding free energy of each of these randomly generated duplexes, we determine their lifetime as explained in the main text;
iv. we sum the duplexes lifetimes and we keep generating random sequences to be duplex partners to the original one until the total lifetime reaches the desired value, chosen to approximate the characteristic experimental time (in this work we assumed $\tau_{TOT} = 10$ h);
v. we generate a new original sequence and repeat the procedure.

This approach enables us to evaluate the population of double helices that are expected after a given time from the preparation of the ranDNA solution and its cooling at room temperature. The resulting distributions are shown in Fig. 4*C* of the main text, where it can be appreciated that for the shortest sequence (6*N*) this kinetic approach leads to a distribution identical to the calculated equilibrium distribution. 12*N* is also almost equilibrated, whereas 20*N* appears far from equilibrium.

Nonequilibrium distributions were evaluated on the basis of about 50 original sequences.

Comparison Between Hybridization in Genomic PCR and Duplex Formation in 20*N*.

This section is devoted to comparing the observed and modeled behavior of 20*N* with the selective hybridization exploited in genomic polymerase chain reaction (PCR). For the sake of the comparison, we will assume PCR primers to be also 20-mers, even though the conclusions will not depend on this assumption.

In PCR, primers efficiently bind to their target segment on longer DNA strands. We consider here genomic PCR because this is the case where the DNA template strands are the longest. Again, for the sake of a clear comparison, we select here a representative example of genomic PCR, performed according to the following protocol. A total mass of 200 ng of mouse genomic DNA, whose size is about 2.5×10^9 base pairs, are dissolved in 50 μL of buffered water. In the same solution, 10^{-12} mol of each of the two primers are dissolved. Annealing of the primers to their target sequence is achieved by cooling the solution, previously heated to denaturating T, at 55 °C for 30 sec.

The efficiency of this procedure contrasts with our observations that 20*N* is kinetically trapped in defected duplex structures.

The two situations have marked quantitative differences: (*i*) the concentration of primers in PCR is of the order of 1.5 μg/mL, each given 20-mer sequence in the 20*N* solutions at the concentrations here considered is over a million times more dilute; (*ii*) the number of potential targets in the DNA template is of order 10^9 for genomic PCR (i.e., considering all the possible positions in which the primer could dock), whereas it is of order 10^{13} in 20*N*. This last figure is obtained by considering the various shifted positions in which each given sequence may produce duplexes with the 10^{12} sequences in the 20*N* solution.

The two situations would become comparable if one could reduce by a factor 10^4 the number of sequences in the 20*N* system while preserving the presence of the perfect complementary sequence for each. This operation would also correspond to dividing $n(\Delta G)$—and hence $P(\Delta G)$—by a factor 10^4 everywhere except for $\Delta G = 19\Delta G_0$, where the current value would be maintained. Such a distribution would be strongly peaked at the energy of perfect pairing, the situation actually exploited in PCR.

A further relevant difference between the two systems is of a kinetic nature. Given the concentration reported above, the mean distance between primers in PCR is about 0.2 μm. Therefore the distance between the target sequence and the nearest primer cannot be larger than that. In 20*N*, instead, the distance between fully complementary 20-mers is about 30 μm. This fact combines with the much larger DNA concentration in the 20*N* system to make the diffusion time spent by each sequence to reach its perfect complementary one many orders of magnitude larger in 20*N* than in PCR even in the absence of intermediate partial pairing and kinetic arrest.

1. Onsager, L (1949) The effects of shape on the interaction of colloidal particles. *Ann NY Acad Sci* 51:627–659.
2. Bolhuis P, Frenkel DJ (1997) Tracing the phase boundaries of hard spherocylinders. *Chem Phys*, 106:666–687.
3. Lu X, Kindt JT (2004) Monte Carlo simulation of the self-assembly and phase behavior of semiflexible equilibrium polymers. *J Chem Phys* 120:10328–10338.
4. Van der Schoot P, Cates ME (1994) Growth static light scattering, and spontaneous ordering of rod-like micelles. *Langmuir* 10:670–679.
5. Bellini T, Cerbino R, Zanchetta G (2011) DNA-based soft phases. *Top Curr Chem*, 10 .1007/128_2011_230.
6. SantaLucia J, Hicks D (2004) The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* 33:415–440.

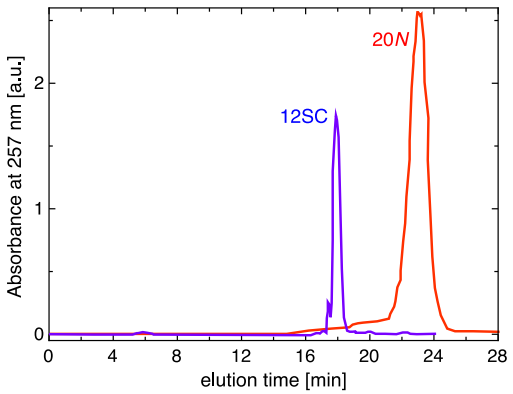

Fig. S1. HPLC traces for the synthesized compound 12SC and 20*N*.

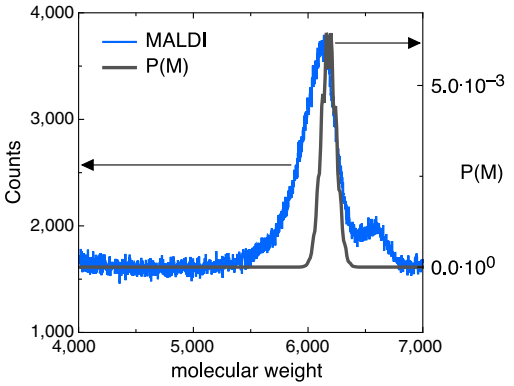

Fig. S2. MALDI-TOF mass spectrum for the 20*N* (blue curve, left axis) compared with the calculated normalized mass distribution $P(M)$ (black line, right axis).

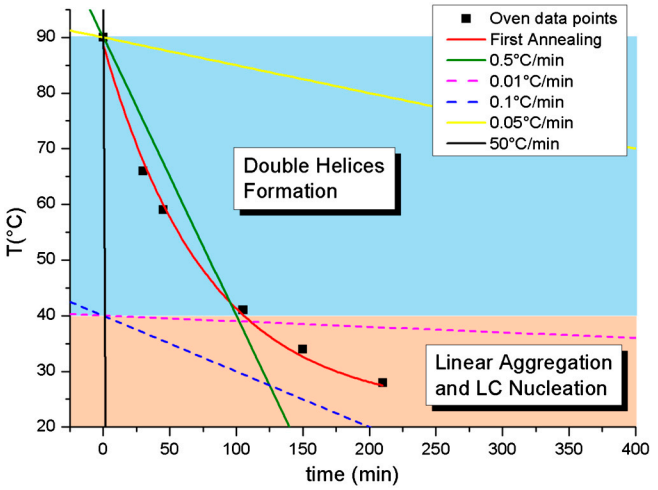

Fig. S3. Schematic summary of cooling rates explored to check for the thermal history dependence of the LC properties. The black squares and red line describe the temperature drop during the first annealing after cell preparation, common to all cells. Shades indicate the two *T* regions, 90–40 and 40–25 °C, relevant for the formation of helical structures (blue shade) and LC development (orange shade).

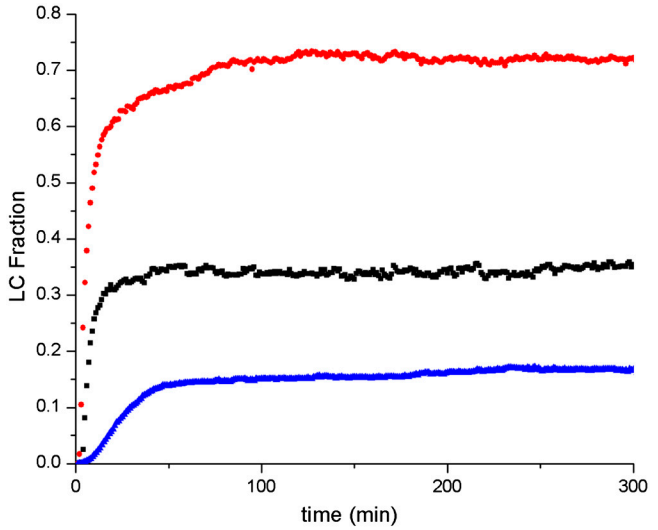

Fig. S4. Time evolution of LC volume fraction at 25 °C for various DNA concentrations (blue triangles, 570 mg/mL; black squares, 690 mg/mL; red circles, 760 mg/mL).

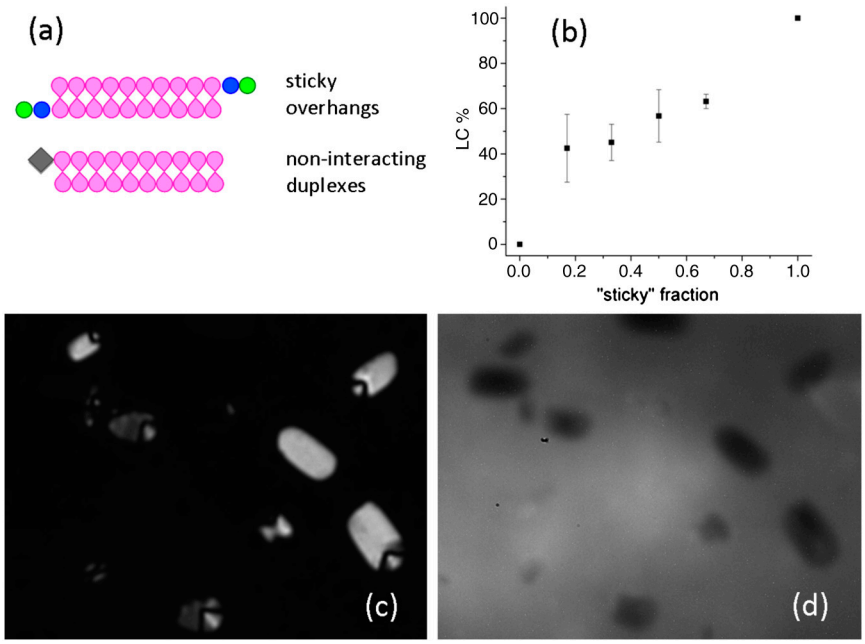

Fig. S5. Phase behavior of mixtures of interacting and noninteracting duplexes. (*A*) Sketch of the two types of helices present in the solutions: a population with sticky unpaired strands is mixed to helices with terminals impairing linear aggregation. (*B*) Average LC fraction as a function of the interacting/non-interacting ratio for set B; (*C* and *D*) DTLM and fluorescence images of a sample from set A: the birefringent domains in *C* are depleted of the fluorescent, noninteracting helices.

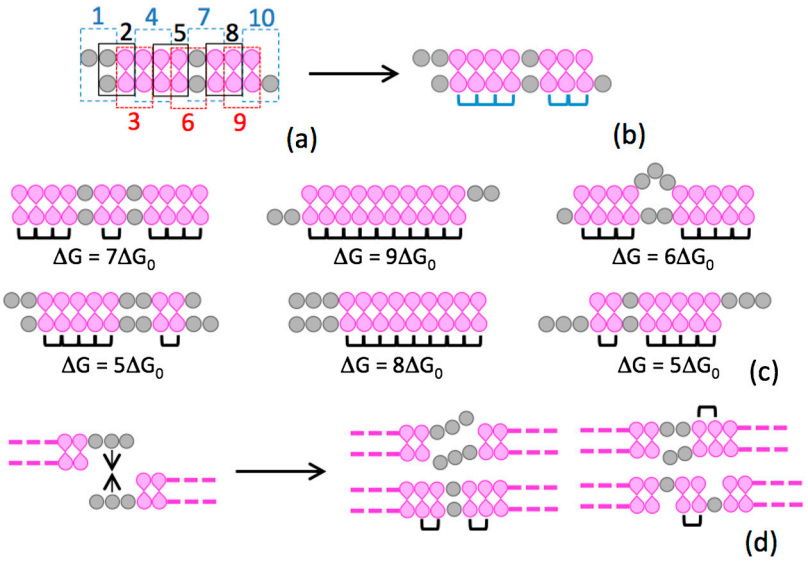

Fig. S6. (*A*) Graphic description of the approach used to evaluate the binding free energy ΔG of hybridized DNA strands relies according to the nearest neighbor model. The double helix is decomposed in quadruplets formed by two consecutive nucleotides on one strand and the corresponding nucleotides on the other strand. Notice that the quadruplets overlap. (*B*) Out of the 10 quadruplets in the example reported in *A*, only five are correctly paired (marked with blue horizontal squared brackets below the drawing). (*C*) Example of the application of the simplified NN approach to compute the energy of variously paired 10-mers. (*D*) Graphical description of the interduplex interactions mediated by random overhangs. The simplified NN model enables evaluating the interduplex energy for overhangs with various complementarity.

Other Supporting Information Files

Table S1 (DOC)
Table S2 (DOC)