Effect of selective end-to-end base stacking interactions on the stability of smectic liquid crystal ordering in concentrated "gapped" DNA solutions

Sineth G. Kodikara[a], Prabesh Gyawali[a], James T. Gleeson [a], Antal Jakli [a,b], Samuel Sprunt[a,b,*], and Hamza Balci[a,*]
[a] Department of Physics, Kent State University, Kent, OH 44242, USA;
[b] Advanced Materials and Liquid Crystals Institute, Kent State University, Kent, OH 44242.

* To whom correspondence should be addressed: hbalci@kent.edu or ssprunt@kent.edu

KEYWORDS: End-to-end stacking, Base-pair energies, Smectic liquid crystal, Gapped DNA, Small-angle x-ray scattering

ABSTRACT:

At sufficiently high concentrations, gapped DNA (GDNA) constructs – two duplex arms connected by a flexible single stranded linker – form layered (smectic) liquid crystalline (LC) phases, whose nature and stability depend on both end-to-end stacking and side-by-side electrostatic interactions between the duplex arms. We performed synchrotron small angle x-ray scattering (SAXS) experiments over 5-65 °C on GDNA constructs that differ only by the terminal base-pairs at the blunt ends of the duplex arms. Two smectic phases are generally observed at the DNA concentrations studied: a bilayer phase at lower temperature, where the duplexes are stacked end-to-end and also positionally ordered within the layers (smectic-B phase), and a monolayer phase at higher temperature with single duplex layer spacing and short-range order within layers (smectic-A phase). Our key finding is that the stability of the bilayer phase varies remarkably among constructs that differ only in their terminal base-pairs. Bilayers formed by constructs with GC termination at both blunt duplex ends melt into the monolayer phase at temperatures up to 30 °C higher than for those with AT termination, while mixed (AT/GC) terminations have intermediate stability. Correlating this compositional variation with macroscopic effects on the thermal stability of the bilayer smectic phase enables us to determine the relative strength of base pair specific, end-to-end stacking interactions. Modeling the bilayer melting in terms of a temperature-dependent reduction in the average fraction of end-to-end paired duplexes enables us to estimate stacking free energies in solutions of duplexes at physiologically relevant DNA concentrations.

The local DNA concentration in chromosomes, viral capsids, and sperm nuclei is on the order of hundreds of mg/ml,[1] concentrations at which DNA can form highly ordered lyotropic liquid crystalline (LC) phases. *In vitro*, short double-stranded DNA (dsDNA) fragments (~6-20 bp, significantly shorter than persistence length of dsDNA) also exhibit LC phases, including chiral nematic and columnar mesophases.[2–8]

The introduction of a flexible single-stranded DNA spacer ("gap") between two rod-like duplexes, producing a "gapped" DNA (GDNA) construct, has been the key for observing elementary layered (smectic) phases in concentrated DNA solutions.[9,10] In addition to the DNA concentration (cDNA), the layer structure is sensitive to gap length and temperature.[9–11] When the gap length is ≥ 10 nucleotides (nt), a bilayer structure composed of stacked duplexes prevails at ambient temperatures. The spacing between the bilayers slightly exceeds two duplex lengths, suggesting that attractive enthalpic interactions between blunt duplex ends and a net conformational entropy gain from segregating flexible "gap" segments act together to stabilize the bilayer stacking. At higher temperature, or for shorter gaps (~3-9 nt), a monolayer smectic phase (approximately single duplex layer spacing) is observed. No smectic layer structure is observed in concentrated solutions when the gap length is ≤ 2 nt.[10]

For constructs with 10 nt or longer gaps, the positional order of the duplexes within the bilayers is short range (corresponding to a smectic-A LC phase) for DNA concentration in the range 230-260 mg/ml. At higher concentration (260-290 mg/ml) and at ambient temperature, the duplexes pack laterally in a hexagonal array within the bilayers (a smectic-B phase).[11] Thus, at sufficient concentration, gapped constructs exhibit lateral positional order similar to the columnar phase formed by their fully paired analogs (FDNA), while also exhibiting the periodic layer structure characteristic of a smectic (unlike FDNA). The side-to-side interactions between duplexes that presumably stabilize the lateral order strongly depend on ionic content of the solution, a natural consequence of the high negative charge density of the DNA backbone, and are particularly sensitive to the presence of multivalent cations.[12,13] Enhanced computational power and recent advanced in computational modeling of DNA and its interactions with cations and other DNA molecules have enabled realistic all-atom molecular dynamic simulations that probe such interactions at high effective DNA concentrations.[14–16]

The base stacking interactions that help stabilize the bilayer smectic phases in GDNA solutions are of fundamental importance for physiological pathways involving replication, repair, or restructuring of dsDNA. This has motivated measuring the strength of these interactions using different approaches.[17] However, due to experimental limitations, small energies, and the difficulty in disentangling base pairing and base stacking interactions, this fundamental question is still actively investigated. To measure the strengths of these interactions various measurements have been performed, including gel electrophoresis assays on nicked dsDNA, thermal melting UV-Vis spectroscopy measurements on dsDNA constructs with different terminal overhangs, optical tweezers assays on DNA nanobeams with blunt duplex ends (where binding/dissociation of these ends is studied at different applied force), and centrifuge force microscopy assays on stacked DNA

constructs (where dissociation of stacked constructs from the surface at different applied centrifugation speeds is monitored).[18–22] In the current study, we correlate the strength of these interactions with the stability of the bilayer structure in smectic GDNA solutions, which presents the opportunity to measure them in physiologically relevant DNA concentrations.

Figs. 1A-C show the three GDNA constructs used in our study. Each consists of 48 bp duplex arms linked by a 20-thymine single-stranded "gap" segment. The constructs differ only by the terminal base pairs on the blunt ends of the duplex arms: AT or GC on both ends, or AT on one end and GC on the other. We shall refer to these constructs as AT-AT, GC-GC, and AT-GC, respectively. We performed SAXS measurements on samples with ~280 mg/ml DNA concentration on beamline 11-BM at the National Synchrotron Light Source II, with the temperature varied from 5 to 65 °C in ~5 °C steps. The range of scattering wavenumber (q) recorded in the SAXS measurements enabled monitoring the smectic layering (diffraction peaks at q of order the inverse duplex length) simultaneously with the lateral, in-layer order of the duplexes (q of order the inverse duplex diameter). We determine characteristic melting temperatures at which the bilayer structure and the in-layer ordering of the constructs in the smectic-B phase are destabilized. These results establish the central role of attractive end-to-end interactions between duplexes in stabilizing the bilayer phase. We also quantify the enthalpic and entropic contributions to the free energy associated with the bilayer ordering.

RESULTS AND DISCUSSION

A typical SAXS pattern from our GDNA samples at ambient temperature features several sharp, small-angle peaks that represent multiple orders of diffraction due to the stacking of duplexes in smectic layers, along with a sharp peak at a wider angle that results from lateral positional ordering of the duplexes. In Fig. 2, azimuthally averaged SAXS intensity (on a log scale) vs q are plotted at various temperatures for the studied GDNA constructs. Four orders of small-angle diffractions are clearly detected. The wave number of the fundamental order (q_1) corresponds to spatial periodicity $d = \frac{2\pi}{q_1} = 34 - 35$ nm at 7 °C. The values of d are slightly greater than the 32.6 nm length of two 48-bp duplexes, and are consistent with a bilayer smectic structure.[10,11] In addition, a sharp peak is recorded at wider angle. The position of this peak is consistent with the first order diffraction (wavenumber q_w) from hexagonal lateral packing of the duplexes in the smectic-B phase of GDNA constructs previously reported at similar concentration and temperature.[11] The corresponding center to center duplex spacings are $a = 4\pi/\sqrt{3}q_w = 3.3 - 3.4$ nm at 7 °C.

Fig. 2A shows temperature-dependent intensity profiles vs q of x-rays diffracted at different angles for the AT-AT GDNA construct. The odd order (first and third order) peaks at q_1 and $q_3 \approx 3q_1$ completely melt along with the wider-angle peak (q_w) at ~43 °C while the even order peaks (second and fourth order peaks at q_2 and $q_4 \approx 2q_2$) survive up to (at least) 65 °C, the highest temperature studied, with a slight upward shift. The peaks at q_1, q_3, and q_w are therefore associated purely with the bilayer smectic-B phase. Their disappearance at ~43 °C indicates a transition to a

"monolayer" smectic-A phase, with layer spacing slightly greater than a single duplex and liquid-like correlations within the layers (accounting for the broad peak centered around a q value slightly below q_w).

The intensity profiles for the AT-GC and GC-GC samples (Figs. 2B and 2C) evolve similarly with increasing temperature; however, the smectic-B to A transition temperatures are significantly higher than for the AT-AT sample. The bilayer peaks and wider-angle peak in the AT-GC sample are detectable up to ~57 °C, while in the GC-GC sample these peaks (and hence the smectic-B layer structure) survive (at least) up to 65 °C. Evidently, the thermal stability of the bilayer smectic B phase is lowest for AT-AT termination on the duplex ends, highest for GC-GC, and intermediate between the two for AT-GC. This difference in bulk behavior among the three systems is due solely to the variation of single base pairs terminating the blunt ends of the duplexes, since otherwise the three GDNA constructs are identical, and the solutions were prepared in the same manner with consistent final component concentrations.

We analyzed the data in Fig. 2 by calculating the sum of the areas under the first and third order small angle peaks (q_1 and q_3) that are specifically associated with the bilayer stacking. As depicted in Fig. 3A, we first fit the data in each peak to a Gaussian function of q, and then integrated the fit result over q (excluding any contribution from the background scattering). Fig. 3B shows the square root of the resulting total integrated intensity under the peaks at q_1 and q_3, normalized to its maximum value and denoted $\sqrt{i_{1+3}}$, as a function of temperature for the three constructs studied. As argued in the Methods section, the normalized quantity $\sqrt{i_{1+3}}$ represents the average fraction (f) of end-to-end "bound" pairs of duplexes making up the GDNA bilayers. The argument assumes that the temperature-dependence of f is proportional to that of the amplitudes ρ_n contributing to the density wave, $\rho = \bar{\rho} + \sum_{n\geq 1} \rho_n \cos q_n z$ (z = axis normal to the layers, $\bar{\rho}$ = average density), that describes the bilayer stacking. This assumption is reasonable if attractive end-to-end interaction between duplexes is the essential mechanism driving the bilayer formation. Only the peaks at q_1, q_3, which are associated with ρ_1, ρ_3, are included in the analysis, because at elevated temperature the peaks at q_2, q_4 are partially overlapped by diffraction (at slightly higher q) from the developing domains of the monolayer phase.

For the three systems studied, Fig. 3B reveals that the fraction f decreases within different temperature ranges that correlate with the specific combination of base pairs terminating the duplex segments of a GDNA construct. Notably in the case of the AT-GC construct, the decrease appears to occur in two steps separated by an inflection point ("kink") around 45 °C. The dashed and solid lines are fits of the data in Fig. 3B for $\sqrt{i_{1+3}} = f$ to single and double Hill functions,[23–25] respectively. Expressions for the Hill functions are given in the Methods section; they are conventionally used to model the fraction of (non-covalently) bound pairs of interacting macromolecules in a solution as a function of temperature. Table 1 gives results for the fit parameter T_m (temperature at which half of the blunt duplex ends are stacked on average) from the fits using a single Hill function, and for the two parameters T_{m1} and T_{m2} determined from fits

using a double Hill function. In all three systems studied, the double Hill function clearly provides a better description of the data than the single function, suggesting that the bilayers, formed by end-to-end paired constructs, melt via a "multi-step" process.

We can correlate the values for T_m determined from the Hill function analysis of $\sqrt{i_{1+3}}$ with values of the end-to-end interaction ("stacking") energies, since higher stacking energies result in higher T_m. The various stacking energies (see Methods section for nomenclature used to describe these energies) have been measured recently by single-molecule techniques.[18,22,26] Using the values reported by Kilchherr *et al.*[18] to compute an average stacking energy for each GDNA system studied, we find $|E_{av,\text{AT-AT}} = -1.57| < |E_{av,\text{AT-GC}} = -1.94| < |E_{av,\text{GC-GC}} = -2.37|$ (in units of kcal/mol), which is exactly in line with the sequence of increasing melting temperatures T_m (Table 1) obtained from the Hill function fits to the data in Fig. 3. Here, $E_{av,\text{XY-X'Y'}}$ represents the average energy of all stacking conformations that are relevant for our respective constructs. On the basis of this comparison, the key result from the single-molecule studies that GC-GC end-to-end interactions are on average significantly stronger than AT-AT interactions may now be extended to bulk systems of highly concentrated duplexes.

To discuss single versus multi step melting of the GDNA bilayers (and to help visualize the possible duplex stackings described above), it is useful to consider a specific model of a GDNA bilayer. We previously proposed[9–11] that provided the gap segment is sufficiently long, the bilayers consist of folded GDNA constructs with the blunt ends of one construct paired with those of another. Figs. 1D-F show schematically six different ways our GDNA constructs can fold and pair. These can be divided into two groups based on the strength of the associated stacking interactions. Figure 1D illustrates the groups for the AT-AT construct. The purple (cyan) shading distinguishes duplexes with the terminal A in the 5′ (3′) position. In Group 1, the stacking energies for each pair of opposing duplexes are degenerate and equal to $E_{AT:AT}$. In Group 2, the energies are either $E_{TA:AT}$ (purple duplexes stacked against each other) or $E_{AT:TA}$ (cyan stacked against cyan). As indicated in Figs. 1E and F, stacking interactions among paired, folded AT-GC and GC-GC constructs may be similarly grouped.

Again using the results of single-molecule measurements in Kilchherr et al[18], the average stacking energy in a pair of folded AT-AT constructs is $E^{(1)}_{av,AT\text{-}AT} = E_{AT:AT} = -1.36$ kcal/mole for Group 1 (indicated by the superscript 1) and $E^{(2)}_{av,AT\text{-}AT} = (E_{TA:AT} + E_{AT:TA})/2 = -1.68$ kcal/mol for Group 2. Similarly, we get $E^{(1)}_{av,AT\text{-}GC} = E_{TA:CG} = -1.39$ kcal/mol, $E^{(1)}_{av,GC\text{-}GC} = E_{Gc:GC} = -1.64$ kcal/mol and $E^{(2)}_{av,AT\text{-}GC} = (E_{TA:AT} + E_{GC:CG})/2 = -2.22$ kcal/mol, $E^{(2)}_{av,GC\text{-}GC} = (E_{CG:GC} + E_{GC:CG})/2 = -2.74$ kcal/mol for Groups 1 and 2 of the AT-GC and GC-GC constructs. We can then argue that the bilayers in each case melt in two steps characterized by temperatures T_{m1}, associated with melting of domains in Group 1 (lower stacking energy cost), and T_{m2}, associated with melting of Group 2 domains (higher stacking energy cost). As discussed above and evident in Fig. 3B, the corresponding analysis of $\sqrt{i_{1+3}}$ with two Hill functions describes the

data well. Moreover, the ascending order of values of T_{m1} and T_{m2} match the ordering of $E_{av}^{(1)}$ and $E_{av}^{(2)}$, respectively, among the three systems studied. However, the argument (and particularly the definition of $E_{av}^{(2)}$) assumes that the discrete pairs of constructs in the two groups shown in Figs. 1D-F have an effective end-end interaction that is a hybrid of the component end-end stackings. If, on the other hand, the bilayer structure is formed from stacking extended ("unfolded") constructs, or if the folded pairs are not discrete (*e.g.*, single ends of one folded construct paired with single ends of two others)[10], this assumption is questionable, and the bilayer melting may instead be determined by some other combination(s) of the non-degenerate end-end interaction energies. In any case, given their limited resolution, the present data are adequately described by an analysis based on a two-step model.

Based upon end-to-end unbinding of stacked duplexes as the mechanism for the melting of the bilayer smectic, we can calculate the Gibbs free energy of stacking interactions. As described in the Methods section, we first obtain the effective equilibrium "binding" constant characterizing the duplex end-end stacking interactions, $k_{eq} = \frac{average\ fraction\ of\ end-end\ "bound"\ duplexes}{average\ fraction\ "unbound"} = \frac{f}{1-f}$, as a function of temperature, using the data for f in Fig. 3. Fig. 4 shows the results for $\ln k_{eq}$ plotted against $1/T$. Fits of these results to the van't Hoff equation[27,28], $\ln k_{eq} = -\frac{\Delta G}{RT}$, yield the following Gibbs free energies evaluated at 296 K: $\Delta G = -0.45 \pm 0.03$ kcal/mol for the AT-AT construct, -1.54 ± 0.16 kcal/mol for the AT-GC construct, and -1.82 ± 0.16 kcal/mol for the GC-GC construct. The entropic and enthalpic contributions to ΔG at 296 K are listed in Table 2. These values reinforce the conclusion that the GC-GC stacking interactions are the most stable, while AT-AT are the least, with AT-GC intermediate between the two.

Free energy measurements using other approaches[18–21] show that stacking of purines with purines is more stable than stacking of pyrimidines with pyrimidines and specifically that GC-GC stacking is more stable than AT-AT stacking, with AT-GC stacking being between these two limits, in agreement with our results. The variation in assay conditions, particularly ionic conditions, and level of decoupling between base-pairing and base-stacking interactions make it challenging to perform a more detailed quantitative comparison. To illustrate, 20 mM $MgCl_2$ or 500 mM NaCl were used in one study,[18] while 50 mM phosphate buffer saline (ionic strength ~110 mM) was used in another,[22] and 10 mM $MgCl_2$ and a few mM KCl were used in yet another study.[26] Using thermal melting studies, Yakovchuk *et al.* characterized variation of stacking free energy with monovalent cation concentration for different base configurations, and demonstrated that the free energies become ~30% more negative (more stable stacking) as the salt concentration was increased from 15 to 100 mM NaCl.[21] They also demonstrated that base-stacking and not base-pairing was primarily responsible for the enhanced stability at higher ionic strength. Furthermore, Kilchherr *et al.*[18] compared the breaking time of their stacked nanobeams in 20 mM $MgCl_2$ to that in 500 mM NaCl and observed that the stacking interactions were more stable in 20 mM $MgCl_2$, confirming the effect of divalent cations in lowering ΔG. In general, the free energies we report

are on the less negative end of the spectrum of values from the studies cited above, which is consistent with the absence of divalent cations in our samples.

Another significant parameter which would impact the stacking free energies is the level of crowding of the environment, which would particularly influence the entropic contributions to the free energy. Most of the recent studies we cite[18,22,26] are single molecule studies where there is effectively zero interaction between neighboring stacked duplexes; however, even typical bulk studies use DNA concentrations that are ~3 orders of magnitude lower than those we use (*e.g.* ~ μM in thermal melting assays[21] *vs.* mM in our studies, which is required for LC phase formation). Our measurements are unique in this respect as they probe base stacking interactions at physiologically relevant DNA concentrations. Relatedly, all free energies we report are based on transition from a bilayer smectic phase to a monolayer smectic phase, which in our model (Fig. 1D–F) results from breaking pairs of stacked duplex arms. Therefore, as described earlier, the free energies we report have inherent averaging, which is not the case in the single molecule approaches.

In order to determine characteristic melting temperatures of the in-layer order that can be compared to the T_m values associated with the bilayer melting, we calculated the area under the sharp, wider-angle peak (q_w) in Fig. 2 in similar fashion to the area under the small angle peaks, as illustrated in Fig. 3A. Fig. 3C shows the results for the square-root of integrated intensity ($\sqrt{i_w}$) vs temperature, together with fits to single and double Hill functions. Values for the parameters T_m or T_{m1} and T_{m2} are listed in Table 1. These values are systematically lower than those for the bilayer peaks for all three GDNA constructs, which indicates that, in heating, the in-layer order starts to melt before end-to-end stacking interactions are broken. Similarly, when cooling down, the duplex ends stack before in-layer order emerges. These findings imply end-to-end stacking interactions stabilize first and promote the lateral ordering of duplexes during formation of the bilayer smectic-B phase.

It should be noted that the measurements reported in this work were performed in the absence of any multivalent cations, which are proposed to enhance attractive interactions between duplexes when present at sufficiently high concentrations. Even though the enhancement is expected to be negligible at physiologically relevant ion concentrations,[29] it would be of fundamental interest to investigate the impact of multivalent cations on the stability of both bilayer and lateral ordering of GDNA duplexes.

CONCLUSIONS

Based on SAXS measurements on concentrated solutions of gapped DNA constructs differing only in terminal base pairs at the blunt ends, we established a correlation between the thermal stability of the bilayer smectic-B phase in GDNA and the relative strength of the duplex end-to-end stacking interactions. Lateral order of the duplexes consistently onsets at a lower temperature than their end-to-end stacking. GC-GC terminated constructs produce the most stable bilayers, while the AT-AT constructs yield the least stable. AT-GC constructs form bilayers with intermediate stability. We also presented evidence of a multi-step bilayer melting process, which

we rationalized using a model of bilayers composed of folded, end-to-end paired GDNA constructs – a model previously introduced to describe elementary smectic ordering of GDNA with sufficiently long gap segments.

The present study demonstrates that the relative strengths of base stacking interactions can be studied at physiologically relevant DNA concentration, temperature range, and ionic conditions via the stability of smectic LC phases formed by gapped DNA constructs. In addition to their significance in liquid crystalline phase formation, base stacking interactions are of fundamental biological significance; our results provide important insights for modeling efforts involving these interactions. The elevated stability of the smectic phases, particularly those formed by stacking GC-GC terminated duplexes, also suggests potential use in biosensor applications where physiological temperature range is of primary focus.

METHODS

GDNA Synthesis: This study utilized three different GDNA constructs, all consisting of 48-base pair duplexes bridged by a single-strand gap of 20 Thymine bases (48-20T-48) with either AT or GC base pairs terminating each of the duplex arms (Figs. 1A-C). The GDNA constructs were formed by annealing three PAGE-purified strands (O1, O2, and O3), purchased from ExonanoRNA (Columbus, OH) or GenScript (Piscataway, NJ). The sequences of O1, O2, and O3 are as follows:

O1 Strand: 5′-**N**CAGATGCACATATCGAGGTGGACATCACTTACGCTGAGTACTTCGAA TTTTTTTTTTTTTTTTTTTTAAGCTTCATGAGTCGCATTCACTACAGGTGGAGCTATAC ACGTAGAC**M**
O2 Strand: 5′- TTCGAAGTACTCAGCGTAAGTGATGTCCACCTCGATATGTGCATCTG**K**
O3 Strand: 5′- **R**GTCTACGTGTATAGCTCCACCTGTAGTGAATGCGACTCATGAAGCTT
For AT-AT Construct: **N**=A, **M**=A, **K**=T, and **R**=T
For AT-GC Construct: **N**=A, **M**=G, **K**=T, and **R**=C
For AT-GC Construct: **N**=G, **M**=G, **K**=C, and **R**=C

The GDNA synthesis process, sample loading procedure into the borosilicate X-ray capillaries, and details of the SAXS measurements are described in Supplementary Information and in previous publications[10,11].

Nomenclature for Stacking Energies: The "AT-AT" construct in Fig. 1A has AT as the terminal base pair on both duplex arms, which results in AT-AT type stacking between neighboring GDNA constructs. The "AT-GC" construct (Fig. 1B) has an AT terminal base pair on one of the duplexes and a GC base pair on the other duplex. This results in a mix of stacking interactions where AT base pairs could stack with either AT or GC base pairs, and vice versa. The "GC-GC" construct (Fig. 1C) has GC as the terminal base pair on both duplex arms, which results in GC-GC type stacking.

In the AT-AT system, there are three distinct stacking energies. One, designated $E_{AT:AT}$, corresponds to A's (T's) lining up against A's (T's), with the 5′ (3′) ends of the single strands on one duplex opposing the 3′ (5′) ends on the other. Here (and below) the subscript $XY:X'Y'$ denotes

$3'$-X stacked opposite $5'$-X' and $5'$-Y stacked opposite $3'$-Y'. Two additional stacking energies are associated with A's (T's) lining up against T's (A's): $E_{TA:AT}$ or $E_{AT:TA}$ if the terminal A's are in the $5'$ or $3'$ positions, respectively. Analogously, the distinct stacking energies for the GC-GC system may be designated $E_{GC:GC}$, $E_{CG:GC}$, and $E_{GC:CG}$. Finally, the AT-GC system introduces four additional non-degenerate stacking energies, $E_{AT:GC}$, $E_{AT:CG}$, $E_{TA:GC}$, and $E_{TA:CG}$. However, only the last of these ($E_{TA:CG}$) is relevant in our samples; the others apply when the A and G bases are both in the $5'$ or $3'$ positions, or when A is in the $3'$ and G in the $5'$ position – none of which are the case for our construct (Fig. 1B). Also, in this nomenclature $E_{XY:X'Y'} = E_{Y'X':YX}$, *e.g.*, $E_{TA:CG} = E_{GC:AT}$, as they both describe stacking of the same duplexes where the bases of one or the other duplex are registered first. Therefore, not every possible combination is written explicitly in this description. In addition to $E_{TA:CG}$, stacking interactions with energies $E_{TA:AT}$ or $E_{GC:CG}$ are possible in our AT-GC system.

X-ray scattering intensity from bi- and mono-layer stacking of GDNA constructs: The mass density wave describing the layer structure may be written as $\rho(\vec{r}) = \bar{\rho} + \sum_{n\geq 1} \rho_n \cos q_n z + \sum_{n\geq 1} \rho'_{2n} \cos q'_{2n} z$, where $\bar{\rho}$ is the average density, ρ_n and ρ'_{2n} are real, non-negative Fourier amplitudes for bi- and mono-layer contributions, respectively, to the overall density, z is the direction of the layer normal, $q_n = nq_0$, and $q'_{2n} = 2nq'_0$ with $q'_0 \gtrsim q_0 = 2\pi/d$ (d = bilayer periodicity). (The latter inequality reflects the slight upward shift in the peaks at q'_2, q'_4 observed in the monolayer smectic-A phase compared to q_2, q_4 in the smectic-B phase.) The intensity of X-rays scattered from $\rho(\vec{r})$ is $I(\vec{q}) \propto \int d\vec{r} \int d\vec{r}'\, \rho(\vec{r})\rho(\vec{r}')\exp\left[i\vec{q}\cdot(\vec{r}-\vec{r}')\right]$. Inserting the expression for $\rho(\vec{r})$ and taking $\vec{q} = q\hat{z}$, we obtain a series of delta functions in $q - q_n$, which experimentally are broadened both by the finite size of smectic domains and by limited resolution in q. The finite size effect can be accounted for by replacing the delta function peaks with Gaussian lineshapes centered on each q_n and with widths that scale with the inverse square of the characteristic size $L = Nd$ of the domain along the layering direction (Warren approximation, N = number of bilayers spaced by d).[30] Convolution with a Gaussian resolution function (whose width is estimated from the experimental conditions to be about 1.5 times the spacing between data points in Fig. 2) broadens the lineshape, while maintaining a Gaussian profile. Fitting the peaks at $q_1 = q_0$ and $q_3 = 3q_0$ to Gaussian profiles and integrating the fitted profiles over q gives $I_{1+3} = \int dq I_1(q) + \int dq I_3(q) \propto (\rho_1^2 + \rho_3^2)N$. (A factor of N^2 that comes from the Warren approximation is reduced to N after the integration over q.) Normalizing I_{1+3} to its maximum value at low temperature then yields the quantity i_{1+3} as a function of temperature.

Characteristic Melting Temperature and Free Energy Calculations: The results for i_{1+3} were analyzed assuming that the Fourier amplitudes ρ_1 and ρ_3 (related specifically to bilayer stacking) are proportional the temperature-dependent, average fraction f of end-to-end paired ("bound") duplexes in a typical sample volume, whence $I_{1+3} \propto f^2$. The characteristic number N of bilayers in a finite sized domain could also vary with temperature (T), but we did not to observe any significant change with T of the SAXS peak widths at q_1 and q_3. (Within the experimental resolution in q, the peak widths should scale as N^{-2}.) Normalization to the maximum I_{1+3} gives

$i_{1+3} = \frac{I_{1+3}}{I_{1+3}^{max}} = f^2/f_{max}^2$, and then assuming $f_{max} = 1$ (at sufficiently low temperatures), we get $i_{1+3} = f^2$. Values of $\sqrt{i_{1+3}}$ versus temperature T were fit to single or double Hill functions, which were used to model $f(T)$ – specifically, $f(T) = 1 - \frac{T^n}{T_m^n + T^n}$ or $f(T) = 1 - \frac{1}{2}\left(\frac{T^n}{T_{m1}^n + T^n} + \frac{T^k}{T_{m2}^k + T^k}\right)$, where T_m, or T_{m1} and T_{m2}, are characteristic melting temperatures, and n and k are positive exponents.

The free energies were calculated from the van't Hoff equation.[27,28] Briefly, the Gibbs free energy (ΔG) can be related to entropy (ΔS) and enthalpy (ΔH) as follows: $\Delta G = \Delta H - T\Delta S$, where T is the absolute temperature. ΔG may also be expressed in terms of an equilibrium "binding" constant (k_{eq}) as $\Delta G = -RT \ln k_{eq}$, where $R = 1.987$ cal/mol/K is the universal gas constant and $k_{eq} = \frac{average\ fraction\ of\ end-end\ "bound"\ duplexes}{average\ fraction\ "unbound"} = \frac{f}{1-f}$. Therefore: $\ln k_{eq} = -\frac{\Delta G}{RT} = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}$. We calculated $\ln k_{eq}$ as a function of T from the data for $f = \sqrt{i_{1+3}}$ (Fig. 3) and then plotted $\ln k_{eq}$ as a function of $\frac{1}{T}$ for temperatures in the transition region (Fig. 4). After fitting the result to a line and determining ΔH from the slope and ΔS from the intercept, we calculated ΔG for $T =$ 296 K.

Acknowledgements: The research reported here was supported by the National Science Foundation under grant DMR-1904167. The authors are particularly grateful to Ruipeng Li and Masa Fukuto for their assistance in performing the SAXS/WAXS measurements on the CMS beamline (11-BM) at the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704.

Supporting Information Available: Details of GDNA synthesis and SAXS measurements

REFERENCES

(1) Livolant, F. Ordered Phases of DNA in Vivo and in Vitro. *Phys. A Stat. Mech. its Appl.* **1991**, *176* (1), 117–137. https://doi.org/10.1016/0378-4371(91)90436-G.

(2) 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* (5854), 1276–1279. https://doi.org/10.1126/SCIENCE.1143826.

(3) Livolant F. Condensed Phases of DNA. *Prog Polym Sci* **1996**, *21*, 1115–1164.

(4) Davidson, M. W.; Strzelecka, T. E.; Rill, R. L. Multiple Liquid Crystal Phases at High DNA Concentrations. *Nature* **1988**, *331*, 457–460.

(5) 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* (9), 8508–8516. https://doi.org/10.1021/acsnano.6b03622.

(6) Durand, D.; Doucet, J.; Livolant, F. A Study of the Structure of Highly Concentrated Phases of DNA by X-Ray Diffraction. *J. Phys. II* **1992**, *2* (9), 1769–1783. https://doi.org/10.1051/jp2:1992233.

(7) 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* (4), 1110–1115. https://doi.org/10.1073/pnas.1117463109.

(8) 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* (33), E7658–E7664. https://doi.org/10.1073/pnas.1721369115.

(9) Salamonczyk, M.; Zhang, J.; Portale, G.; Zhu, C.; Kentzinger, E.; Gleeson, J. T. T.; Jakli, A.; De Michele, C.; Dhont, J. K. G. K. G.; Sprunt, S.; Stiakakis, E.; Jakli, A.; Michele, C. De; Sprunt, S.; Stiakakis, E.; De Michele, C.; Dhont, J. K. G. K. G.; Sprunt, S.; Stiakakis, E. Smectic Phase in Suspensions of Gapped DNA Duplexes. *Nat. Commun.* **2016**, *7* (May), 1–9. https://doi.org/10.1038/ncomms13358.

(10) Gyawali, P.; Saha, R.; Smith, G. P.; Salamonczyk, M.; Kharel, P.; Basu, S.; Li, R.; Fukuto, M.; Gleeson, J. T.; Clark, N. A.; Jákli, A.; Balci, H.; Sprunt, S. Mono- and Bilayer Smectic Liquid Crystal Ordering in Dense Solutions of "Gapped" DNA Duplexes. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (12). https://doi.org/10.1073/pnas.2019996118.

(11) Gyawali, P.; Saha, R.; Kodikara, S. G.; Li, R.; Fukuto, M.; Gleeson, J. T.; Smith, G. P.; Clark, N. A.; Jakli, A.; Balci, H.; Sprunt, S. Smectic-B Phase and Temperature-Driven Smectic-B to -A Transition in Concentrated Solutions of "Gapped" DNA. *Phys. Rev. Res.* **2022**, *4* (3), 033046. https://doi.org/10.1103/PHYSREVRESEARCH.4.033046/FIGURES/4/MEDIUM.

(12) Wong, G. C. L.; Pollack, L. Electrostatics of Strongly Charged Biological Polymers: Ion-Mediated Interactions and Self-Organization in Nucleic Acids and Proteins. *Rev. Adv.*

Annu. Rev. Phys. Chem **2010**, *61*, 171–189. https://doi.org/10.1146/annurev.physchem.58.032806.104436.

(13) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding Nucleic Acid-Ion Interactions. *Annu. Rev. Biochem.* **2014**, *83*, 813–841. https://doi.org/10.1146/annurev-biochem-060409-092720.

(14) Maffeo, C.; Luan, B.; Aksimentiev, A. End-to-End Attraction of Duplex DNA. *Nucleic Acids Res.* **2012**, *40* (9), 3812–3821. https://doi.org/10.1093/NAR/GKR1220.

(15) Yoo, J.; Winogradoff, D.; Aksimentiev, A. Molecular Dynamics Simulations of DNA–DNA and DNA–Protein Interactions. *Curr. Opin. Struct. Biol.* **2020**, *64*, 88–96. https://doi.org/10.1016/J.SBI.2020.06.007.

(16) Srivastava, A.; Timsina, R.; Heo, S.; Dewage, S. W.; Kirmizialtin, S.; Qiu, X. Structure-Guided DNA–DNA Attraction Mediated by Divalent Cations. *Nucleic Acids Res.* **2020**, *48* (13), 7018–7026. https://doi.org/10.1093/NAR/GKAA499.

(17) Li, L.; Pabit, S. A.; Lamb, J. S.; Park, H. Y.; Pollack, L. Closing the Lid on DNA End-to-End Stacking Interactions. **2008**. https://doi.org/10.1063/1.2937402.

(18) Kilchherr, F.; Wachauf, C.; Pelz, B.; Rief, M.; Zacharias, M.; Dietz, H. Single-Molecule Dissection of Stacking Forces in DNA. *Science (80-.).* **2016**, *353* (6304). https://doi.org/10.1126/SCIENCE.AAF5508.

(19) Petersheim, M.; Turner, D. H. Base-Stacking and Base-Pairing Contributions to Helix Stability: Thermodynamics of Double-Helix Formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry* **1983**, *22* (2), 256–263. https://doi.org/10.1021/bi00271a004.

(20) Protozanova, E.; Yakovchuk, P.; Frank-Kamenetskii, M. D. Stacked-Unstacked Equilibrium at the Nick Site of DNA. *J. Mol. Biol.* **2004**, *342* (3), 775–785. https://doi.org/10.1016/j.jmb.2004.07.075.

(21) 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* (2), 564–574. https://doi.org/10.1093/nar/gkj454.

(22) Punnoose, J. A.; Thomas, K. J.; Chandrasekaran, A. R.; Vilcapoma, J.; Banco, T.; Halvorsen, K. Quantifying Individual Base Stacking Energies with the Centrifuge Force Microscope Quantifying Individual Base Stacking Energies with the Centrifuge Force Microscope. **2022**, No. May. https://doi.org/10.1101/2022.05.25.493108.

(23) Mergny, J. L.; Lacroix, L. Analysis of Thermal Melting Curves. *Oligonucleotides* **2003**, *13* (6), 515–537. https://doi.org/10.1089/154545703322860825.

(24) Parisiades, P. A Review of the Melting Curves of Transition Metals at High Pressures Using Static Compression Techniques. *Crystals* **2021**, *11* (4). https://doi.org/10.3390/cryst11040416.

(25) Wienken, C. J.; Baaske, P.; Duhr, S.; Braun, D. Thermophoretic Melting Curves Quantify the Conformation and Stability of RNA and DNA. *Nucleic Acids Res.* **2011**, *39* (8). https://doi.org/10.1093/nar/gkr035.

(26) Abhinav Banerjee, A.; Anand, M.; Kalita, S.; Ganji, M. Single-Molecule Analysis of DNA Base-Stacking Energetics Using Patterned DNA Nanostructures. *bioRxiv* **2022**, 2022.09.08.506950. https://doi.org/10.1101/2022.09.08.506950.

(27) van't Hoff, M. J. H. Etudes de Dynamique Chimique. *Recl. des Trav. Chim. des Pays-Bas* **1884**, *3* (10), 333–336. https://doi.org/https://doi.org/10.1002/recl.18840031003.

(28) Laganowsky, A.; Clemmer, D. E.; Russell, D. H. Variable-Temperature Native Mass Spectrometry for Studies of Protein Folding, Stabilities, Assembly, and Molecular Interactions. *Annu. Rev. Biophys.* **2022**, *51* (1), 63–77. https://doi.org/10.1146/annurev-biophys-102221-101121.

(29) Bai, Y.; Das, R.; Millett, I. S.; Herschlag, D.; Doniach, S. Probing Counterion Modulated Repulsion and Attraction between Nucleic Acid Duplexes in Solution. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (4). https://doi.org/10.1073/PNAS.0404448102.

(30) Kaganer, V. M.; Ostrovskii, B. I.; De Jeu, W. H. X-Ray Scattering in Smectic Liquid Crystals: From Ideal- to Real-Structure Effects. *Phys. Rev. A* **1991**, *44* (12), 8158. https://doi.org/10.1103/PhysRevA.44.8158.

	Bilayer Order Melting			In-plane Order Melting		
	Single Hill Function Fit	Double Hill Function Fit		Single Hill Function Fit	Double Hill Function Fit	
	T_m(°C)	T_{m1}(°C)	T_{m2}(°C)	T_m(°C)	T_{m1}(°C)	T_{m2}(°C)
AT-AT	34.8±0.4	29.4±0.3	39.6±0.3	31.4±0.4	25.3±0.4	38.1±0.4
AT-GC	50.0±0.7	42.1±0.7	55.3±0.4	46.8±0.8	37.6±0.8	53.9±0.5
GC-GC	64.4±0.4	51.7±0.2	67.3±0.8	58.6±0.2	47.3±0.8	65.6±0.6

Table 1. Results for the fit parameter T_m from the fits using a single Hill function, and for the two parameters T_{m1} and T_{m2} determined from fits using a double Hill function.

Construct	Intercept (kCal/mol)	Slope (K)	ΔH (kCal/mol)	ΔS (kCal/mol/K)	ΔG (kCal/mol)
AT-AT	65.30±3.04	-19556±913	38.84±1.81	0.13±0.01	0.45±0.03
AT-GC	49.53±3.57	-15435±1130	30.65±2.24	0.10±0.01	1.54±0.16
GC-GC	38.45±2.38	-12296±765	24.42±1.52	0.08±0.01	1.82±0.16

Table 2. Results of the analysis of data presented in Figure 4. The free energies are calculated at 296 K (23 °C).

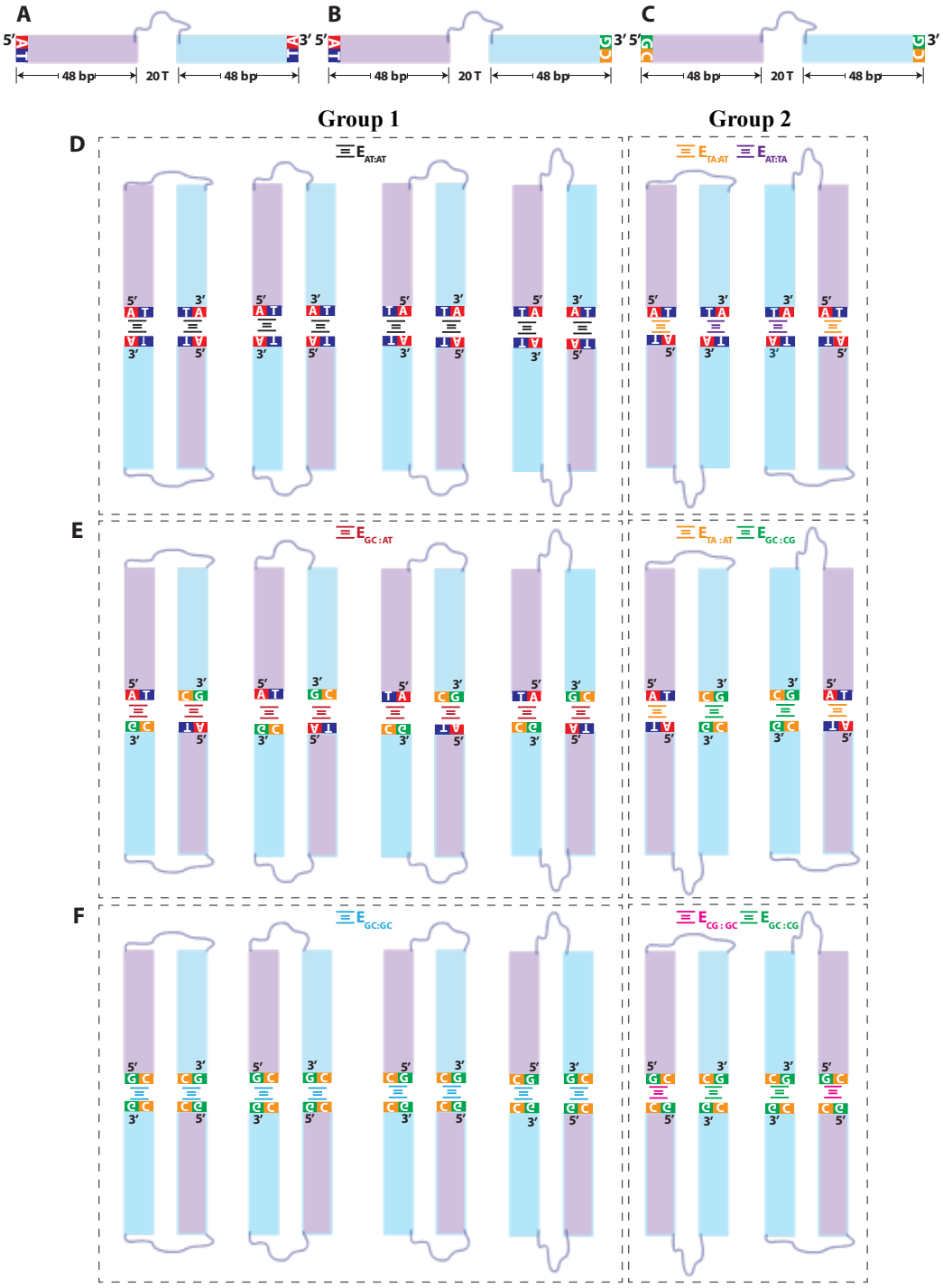

Figure 1: Schematic motifs of 48-20T-48 GDNA constructs. All constructs are identical except the terminal base pairs of the duplex arms. The duplexes are color coded such that duplexes that have an A or a G at 5′ (3′) positions are purple (cyan). (A) AT-AT construct where both duplexes have AT terminal base pairs. (B) AT-GC construct where one of the duplexes has a terminal GC base pair while the other has a terminal AT base pair. (C) GC-GC construct where both duplexes have GC terminal base pairs. (D) — (F) Schematics showing possible end-to-end stacking interactions for folded conformations of the AT-AT construct (D); for AT-GC construct (E); and for GC-GC construct (F). The conformations of each construct are classified as Group 1 and Group 2. The three conformations in Group 1

have degenerate stacking energies, while the two conformations in Group 2 have distinct stacking energies, resulting in a total of three distinct stacking energies for each construct.

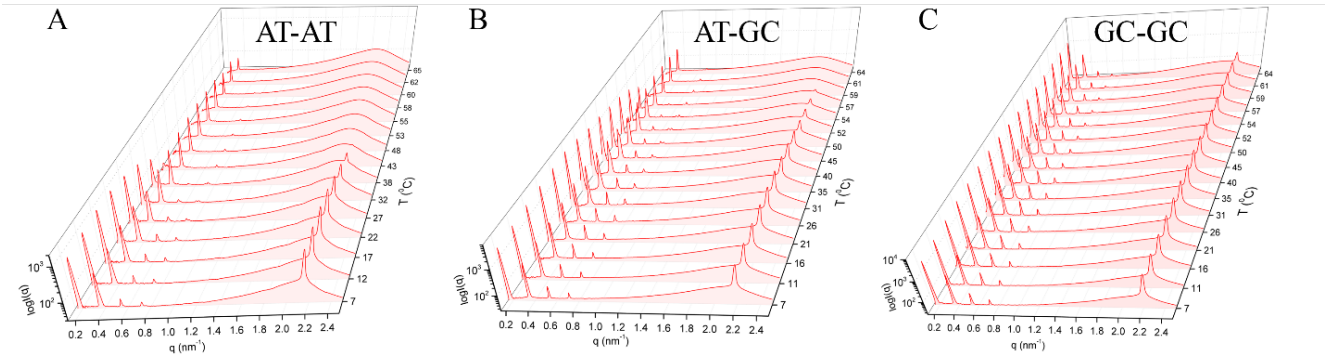

Figure 2: Temperature dependence of the azimuthally averaged SAXS intensity vs. scattering wavenumber q for GDNA samples. The intensity is on a log scale, and the data were acquired on a heating cycle (gradually increasing temperatures). (A) AT-AT construct; (B) AT-GC construct; and (C) GC-GC construct. All three samples are in a bilayer smectic-B phase at low temperatures, which melts at elevated temperature (indicated by diminishment of small-angle peaks at $q_1 \approx 0.18$ nm^{-1} and $q_3 \approx 0.54$ nm^{-1} that are specifically associated with the bilayer spacing and by the disappearance of the sharp wide-angle peak at $q_w \approx 2.2$ nm^{-1}, which is associated with in-layer positional ordering of the duplexes). While the bilayer peaks disappear at about 43 °C and 57 °C for AT-AT and AT-GC constructs, respectively, they do not completely melt for the GC-GC construct (at least up to 65 °C). The bilayer smectic-B melts into a monolayer smectic-A phase with characteristic small-angle peaks at $q_2 \approx 0.36$ nm^{-1} and $q_4 \approx 0.73$ nm^{-1}, and a broad wide-angle peak indicating liquid-like short-range correlations within the layers.

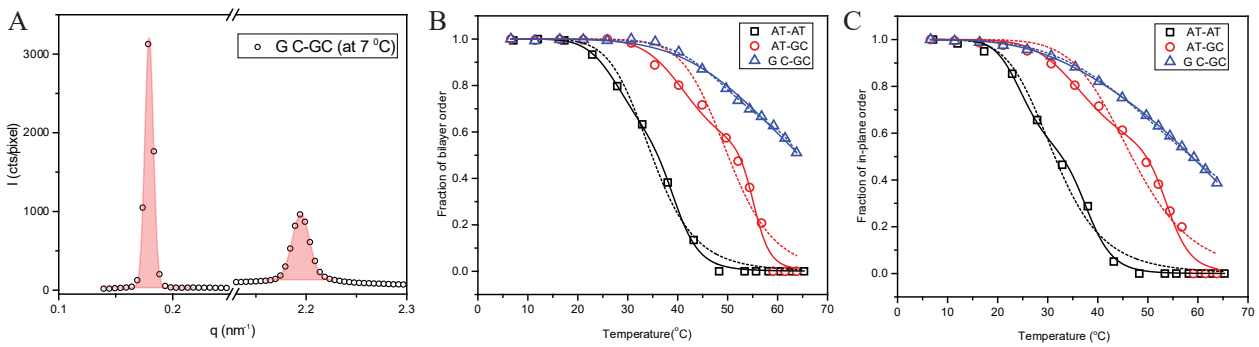

Figure 3: Analysis of the SAXS data in Fig. 2 as a function of temperature. (A) To determine the average fraction (f) of end-end "bound" duplexes constituting the bilayer domains, the areas under Gaussian fits to the small angle peaks at q_1 and q_3 (specifically associated with the bilayer structure) were calculated and summed. The results were normalized to their maximum value, yielding the quantity denoted i_{1+3} (see text). A similar analysis was performed for the sharp wide-angle peak at q_w (associated with in-layer ordering of the duplexes), yielding the quantity denoted i_w in the text. Two representative Gaussian fits are shown for the GC-GC construct at 7 °C for the peaks at q_1 (at left) and q_w (at right). Open symbols are data points and solid lines are the Gaussian fits. The area under each peak is shaded. The base of the Gaussian represents the background. (B) Results for $\sqrt{i_{1+3}}$ plotted vs temperature; as argued in the Methods section, we may associate this quantity with the fraction f. The dashed and solid lines are fits of the data points to a single and double Hill functions, respectively, from which characteristic melting temperatures T_m or T_{m1} and T_{m2} are determined. (C) Results for $\sqrt{i_w}$ vs temperature. The data points were also fit to single and double Hill functions (dashed and solid lines, respectively) in order to obtain characteristic melting temperatures for comparison with those determined from $\sqrt{i_{1+3}}$.

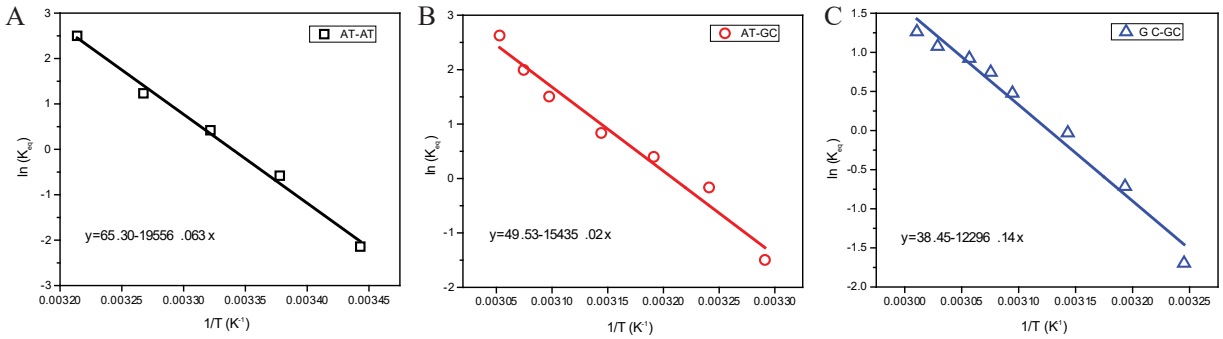

Figure 4: Calculation of Gibbs free energy due to end-to-end stacking of duplexes. Data and linear fit for AT-AT construct are shown in (A); for AT-GC construct in (B); and for GC-GC construct in (C). Symbols are results for the natural logarithm of effective equilibrium "binding" constant (k_{eq}) obtained from the results for $\sqrt{i_{1+3}}$ (see Methods section), and the solid lines are linear fits. The best-fit equation is given in each figure.