Theory of phase segregation in DNA assemblies containing two different base-pair sequence types

You may also be interested in:

Statistical mechanical model for a closed loop plectoneme with weak helix specific forces
Dominic J (O') Lee

Torsional fluctuations in columnar DNA assemblies
D J Lee and A Wynveen

Collapse and coexistence for a molecular braid with an attractive interaction component subject to mechanical forces
Dominic J (O') Lee

The tilt-dependent potential of mean force of a pair of DNA oligomers from all-atom molecular dynamics simulations
Ruggero Cortini, Xiaolin Cheng and Jeremy C Smith

Chiral electrostatics breaks the mirror symmetry of DNA supercoiling
R Cortini, D J Lee and A Kornyshev

Correlation forces between helical macro-ions in the weak coupling limit
D J Lee

Charge renormalization of helical macromolecules
D J Lee

Statistical mechanics of polymers subject to a force
E Orlandini and S G Whittington

The role of electrostatics in the B to A transition of DNA: from solution to assembly
Laura Rudd, Dominic J Lee and Alexei A Kornyshev

New Journal of Physics

The open access journal at the forefront of physics

PAPER

CrossMark

Theory of phase segregation in DNA assemblies containing two different base-pair sequence types

Dominic J (O') Lee[1], **Aaron Wynveen**[2] **and Alexei A Kornyshev**[1]

1   Department of Chemistry, Imperial College London, SW7 2AZ, London, UK
2   School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA

E-mail: domolee@hotmail.com, a.wynveen@googlemail.com and a.kornyshev@imperial.ac.uk

Abstract

Spontaneous pairing of homologous DNA sequences—a challenging subject in molecular biophysics, often referred to as 'homology recognition'—has been observed *in vitro* for several DNA systems. One of these experiments involved liquid crystalline quasi-columnar phases formed by a mixture of two kinds of double stranded DNA oligomer. Both oligomer types were of the same length and identical stoichiometric base-pair composition, but the base-pairs followed a different order. Phase segregation of the two DNA types was observed in the experiments, with the formation of boundaries between domains rich in molecules of one type (order) of base pair sequence. We formulate here a modified 'X–Y model' for phase segregation in such assemblies, obtain approximate solutions of the model, compare analytical results to Monte Carlo simulations, and rationalise past experimental observations. This study, furthermore, reveals the factors that affect the degree of segregation. Such information could be used in planning new versions of similar segregation experiments, needed for deepening our understanding of forces that might be involved, e.g., in gene–gene recognition.

1. Introduction

In this article we develop a statistical mechanical model of DNA assemblies containing mixtures of fragments of two different base-pair sequences that are allowed to phase segregate. This model, e.g., may describe phase segregation of DNA in cholesteric spherulites. Spherulites are liquid crystal structures formed in DNA-non-condensing electrolytic solutions upon applying mild osmotic stress that helps to overcome net repulsion between the DNA molecules. This current study was initially motivated by experiments [1] in which partial segregation of DNA molecules, according to their base-pair sequence texts (i.e., the order of their nucleic bases), was observed in spherulites containing a mixture of two DNA species with different sequence texts.

These experiments were performed by mixing an equal amount of 294 bp-long fragments of synthetic dsDNA with two different sequences in 0.5 M NaCl under mild osmotic stress. Each of the two types of sequences contained the same numbers of GC and AT base pairs, but the base pairs were shuffled randomly into two different orderings. As osmotic pressure was applied slowly, causing the molecules to condense into spherulites, the two sequence types simultaneously segregated, which was observed using confocal microscopy. The fragments were observed to be in a cholesteric arrangement with a large pitch (a small tilt angle between cholesteric layers), and so the arrangement of molecules could be described as being, roughly, in a columnar phase, in which DNA fragments are arranged parallel to each other. To detect segregation between the two species in these experiments [1], each species was labelled by chromophores of different colour (red or green). In order not to overload the samples with chromophores, only 5% of DNA fragments were thus labelled. But this fraction was sufficient to see a signature of homology segregation, in which DNA fragments with the same base pair ordering preferentially condensed in domains richer in their proportion. This was indicated by the appearance of red and green patches in the observed images for each spherulite. The degree of colour segregation

was significant enough that it could not be explained simply by random statistics. Conversely, control experiments in which dsDNA fragments of only one species, half labelled with one colour chromophore and half with the other, demonstrated random mixing of the molecules labelled with the two chromophore types. These ruled out segregation effects due to the type of chromophore, and, hence, a difference in base-pair sequence was shown as the primary driving force behind the segregation.

Such evidence of homology recognition in a protein-free environment has been widely discussed in the scientific press [2] as being an innate consequence of the structure of DNA. Signatures of homology recognition were also reported earlier in gel electrophoresis studies [3]; and some time later, strong indications of the same effect were obtained by single molecule manipulation techniques [4] and AFM studies for DNA on nucleosomes [5]. Further corroborative evidence of homology recognition in a protein free environment was presented in a very recent study, in which preferential interaction of long homologous tracts, in parallel orientation within single molecules, was observed [6] in magnetic bead distance-force experiments.
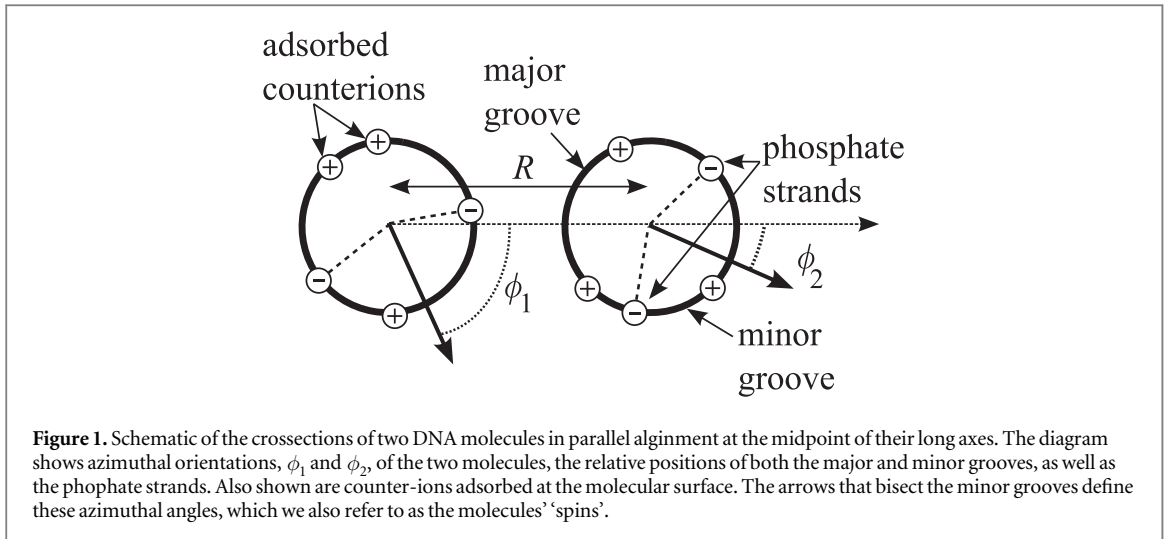
The existence of such interaction forces that distinguish between similar and different molecules might be biologically important. Indeed the pairing of chromosomes of similar sequences is an essential step in many biological processes [7–9]. It has been speculated whether this initial pairing might be some innate general characteristic of DNA [9], resulting, again, from the contrasting forces between DNA with differences in sequence.

Prior to the work of [1], it was shown that the phenomena of phase segregation between different DNA sequences might indeed be theoretically possible. Indeed, an explanation of the results of osmotic stress experiments presented in [1] relies on the idea that the overall interaction energy between each fragment, although net repulsive, should contain a *helix-specific* attractive component. The degree of attraction may be modelled quantitatively through the Kornyshev–Leikin (KL) theory [10, 11]. In the theory, the attractive component is found to be stronger between homologous molecules than between two non-homologous fragments upon considering the non-ideal helix structure of DNA molecules [12, 13]. This difference, referred to as the 'recognition energy', was first calculated in [14]. The theory of homology recognition was then developed with a higher level of sophistication in a series of subsequent works (for review see [10, 15], and for a recent study [16] with citations to later articles contained therein). The recognition energy can comprise several $k_B T$ per persistence length, and it increases with the length of the molecules. However, in the segregation experiments of a liquid-crystal configuration [1], the employed molecules had to be sufficiently short—not much longer than the dsDNA bending thermal persistence length—to form well-ordered liquid crystalline phases. Those molecules used in [1] were roughly twice the persistence length, and according to theory, segregation effects would have been weak. This, indeed, was observed, but distinguishably seen.

Later works have suggested that the preferred conformation of DNA with the same sequence will be parallel (i.e. parallel alignment with base pair sequences running in the same direction) to each other but not antiparallel (parallel alignment but base pair sequences in opposite directions), as this recognition depends on the homologous sequences being in face-to-face register, with base pairs of the same type for the two molecules being directly across from each other [16].

Alternately, preferential pairing between identical DNA sequences might be, in some part, due to interactions that depend locally on the type of base pair content [6, 17, 18]. Here, the interaction energy between two like base pairs has lower interaction energy than two non-alike ones. One such mechanism, mediated by counter-ions being localised by particular base pairs [19], has been discussed in [6]. An alternate mechanism utilising secondary hydrogen bonding was proposed in [17]. Sequence specific variations in van der Waals interactions might also be responsible [18]. Such mechanisms would also be influenced by the base-pair specific pattern of DNA distortions and are manifestly helix-structure-dependent forces. Though, in this work, we assume that the global pairing mechanism due to helix distortions (discussed in the previous paragraph) is the root cause of segregation.

In this paper, we lay the groundwork in developing a model that may rationalise the degree of segregation under different conditions. In the main text we develop a statistical mechanical description taking into account key details of the putative helix specific forces in DNA–DNA interactions; in the supplementary material we append a simpler phenomenological model which we will refer to in the main text for comparison. In fact, the final form of the partition function that we use in our calculations (see the end of next section) is independent of a specific pairing mechanism, up to the choice of model parameters, as it simply relies on the fact that there are helix-structure-dependent forces.

Figure 1. Schematic of the crossections of two DNA molecules in parallel alginment at the midpoint of their long axes. The diagram shows azimuthal orientations, ϕ_1 and ϕ_2, of the two molecules, the relative positions of both the major and minor grooves, as well as the phophate strands. Also shown are counter-ions adsorbed at the molecular surface. The arrows that bisect the minor grooves define these azimuthal angles, which we also refer to as the molecules' 'spins'.

2. Model

2.1. General considerations

Below, we formulate the principles of the theory that could describe phase segregation between two types of DNA fragments, each with differently ordered base pair texts, in liquid crystalline columnar assemblies.

Our theory takes into account thermal fluctuation modes corresponding to rotations of the molecules about their long axes. These modes are important if the helix structure of DNA molecules in DNA–DNA interactions matters. Indeed, for interactions that depend on helix structure, the pair interaction between molecules depends on the relative azimuthal orientation of two molecules about their long axes [10]. Thus, such fluctuations are restricted by helix dependent interactions.

Any pair potential (for two molecules labelled 1 and 2) that depends on the continuous helical symmetry of the molecule, for molecules aligned parallel to each other, may be expanded as a Fourier series in $\phi_1 - \phi_2$

$$V_{1,2}(R) = L \sum_{n=0}^{\infty} a_n(R) \cos(n(\phi_1 - \phi_2)), \tag{2.1}$$

where ϕ_1 and ϕ_2 are the ('spin') angles that both helices (at the midpoint of the long axes of the molecules) make to an axis that lies in a plane perpendicular to the long axes of both the molecules (see figure 1). We use the term 'spin' to emphasise an isomorphism (or analogy) between this approach for modelling interactions in columnar DNA assemblies and a kind of 2D X–Y model (modified to account of an extra cosine(s)) of magnetism. In the latter, the angle describes the orientation of real, magnetic spins. In equation (2.1), R is the inter-axial separation between the two molecules. Also, in writing equation (2.1), the interaction has been assumed to scale linearly with the length of the molecules, L. This supposes the two molecules are fixed to be in complete parallel juxtaposition with each other. In a 2D-assembly of N molecules, there are N such spin angles ϕ_μ describing these degrees of freedom and $N(N-1)/2$ pair potentials.

Our model describes a single layer of DNA molecules, in columnar alignment, whose molecular centres are assumed to be fixed to a 2D regular lattice structure, in a plane perpendicular to the molecular centre lines, as in [20]. Though, as in [20], we allow for the lattice to distort from a hexagonal to a rhombic structure to minimise the free energy in response to the DNA–DNA interactions. As in [20], as well as summing up the pair potentials for the lattice, we truncate the sum in equation (2.1) at $n = 2$. Thus, our theory involves a modified version of an X–Y magnetic model, originally studied in [20].

Reference [20] solely dealt with columnar aggregates comprising only of one type of DNA fragments, i.e., all of them having the same sequence of base pairs. However, here, with the intention of describing the experiments of [1], we extend this model to the case where there are now DNA molecules with two different sequence texts. For there to be phase segregation between molecules with different sequences, there must be a more attractive interaction between like (homologous) molecules than between different (non-homologous) ones, as was suggested by previous works [14–16, 21]. Here, the difference in the interaction between homologous and non-homologous molecules results from the helical distortions, specifically in the twist and rise between adjacent base-pairs, of the DNA double helix. The pattern of distortions is dependent on base pair text [22]. Molecules with the same base pair sequence share the same distortions, and therefore can maintain an energetically preferential charge alignment along their full length, whereas non-homologous molecules, with different sequences and, therefore, different distortions, cannot maintain this alignment. These patterns of distortions

influence the positions of the negatively charged helical phosphate groups along the molecule as well as the positions of the groove centres, where positively charged counter-ions may localise. Therefore, homologous pairs may possess a lower interaction energy than their non-homologous counterparts. This was shown to be the case with the mean-field electrostatic KL-theory [14, 15].

To model helical distortions of the molecules, modifying the model considered previously [20], we replace the constant azimuthal (or, again, 'spin') angle ϕ_μ with $\phi_\mu(z)$ (for a molecule at site μ in the assembly), since the 'spin' angle depends on the position z along the long axes of the molecules due to the distortions of the helical twist of the molecules. (A discussion and justification for this can be found in [10, 14, 15].) Hence, the function $\phi_\mu(z)$ depends on the base pair sequence of the molecule.

The form for the interaction energy is then obtained by replacing L in equation (2.1) with an integral over z along the length of the molecule. As there are DNA molecules of two different sequences, and thus two different distortion patterns, the azimuthal orientation for a molecule at the lattice site μ are labelled either as $\phi_\mu^A(z)$ or $\phi_\mu^B(z)$; with the molecule types A and B. An additional degree of freedom in the system is thus manifested since we allow for the assembly to exchange molecules of type A and B with an external reservoir, and thus allow for the relative proportions of the two types of molecule in the assembly to vary.

Summing up all the pair potentials, we can write the total interaction energy for the assembly (including all pair interactions) as

$$
E_{\text{int}} = \frac{1}{2} \int_{-L/2}^{L/2} \mathrm{d}z \sum_{\mu \neq \nu} \Big[a_0(|\mathbf{R}_\mu - \mathbf{R}_\nu|) - a_1(|\mathbf{R}_\mu - \mathbf{R}_\nu|) \cos(\phi_\mu^\alpha(z) - \phi_\nu^\beta(z))
$$
$$
+ a_2(|\mathbf{R}_\mu - \mathbf{R}_\nu|) \cos(2\phi_\mu^\alpha(z) - 2\phi_\nu^\beta(z)) \Big]. \tag{2.2}
$$

As mentioned, in this expression, we have truncated the sum over helical harmonics (equation (2.1)) at $n = 2$ (as in [20]). The position vectors $\mathbf{R}_\mu$ and $\mathbf{R}_\nu$ are restricted to fixed lattice sites μ and ν. Here, the indices α and β can either be set to A or B, depending on which of the two species of molecules occupies the sites μ and ν, respectively.

Now, the azimuthal angles $\phi_\mu^\alpha(z)$ (and $\phi_\nu^\beta(z)$) can be expressed as

$$
\phi_\mu^\alpha(z) = \phi_\mu^0 + \frac{1}{h} \int_0^z \delta\Omega^\alpha(z') \mathrm{d}z', \tag{2.3}
$$

where, h ($\simeq 3.4$ Å) is the average spacing between two adjacent base-pairs along a molecule. Here ϕ_μ^0 is the azimuthal angle that the DNA double helix associated with the lattice site μ at its molecular centre makes with a single axis (defined throughout the assembly); the function $\delta\Omega^\alpha(z)$ represents the pattern of distortions away from an ideal helix, which, again, is dependent on the base pair structure. For an ideal helix $\delta\Omega^\alpha(z) = 0$. If two neighbouring molecules possess random sequences with respect to each other, i.e., are non-homologous, the two functions $\delta\Omega^\alpha(z)$ are uncorrelated over large length scales along the DNA molecules, and therefore obey Gaussian statistics:

$$
\langle \delta\Omega^A(z) \rangle_\Omega = \langle \delta\Omega^B(z) \rangle_\Omega = 0,
$$
$$
\langle \delta\Omega^A(z) \delta\Omega^A(z') \rangle_\Omega = \langle \delta\Omega^B(z) \delta\Omega^B(z') \rangle_\Omega = \frac{h^2}{\lambda_c^{(0)}} \delta(z - z'),
$$
$$
\langle \delta\Omega^A(z) \delta\Omega^B(z') \rangle_\Omega = \langle \delta\Omega^B(z) \delta\Omega^A(z') \rangle_\Omega = 0. \tag{2.4}
$$

In equation (2.4), the subscript Ω of the averaging brackets refers to ensemble averaging over all possible base pair realisations.

The parameter $\lambda_c^{(0)} = (1/\lambda_{\Omega,\Omega} + 1/\lambda_{h,h} - 2/\lambda_{\Omega,h})^{-1}$ is the (intrinsic) helical coherence length [22] of DNA molecules, where $\lambda_{\Omega,\Omega}$ and $\lambda_{h,h}$ are the contributions due to variations from the mean values of, respectively, twist angles and base pair rises (vertical distances) between adjacent base-pairs along any DNA molecule. The correlation length $\lambda_{\Omega,h}$ arises from a correction due to correlations between the rise and the twist angle. One can represent $\lambda_{\Omega,\Omega} = h/\langle \delta\Omega^2 \rangle$ and $\lambda_{h,h} = h/(g^2 \langle \delta h^2 \rangle)$ (here $g = 2\pi/H$, where H is the average helical pitch of DNA), where $\langle \delta\Omega^2 \rangle$ and $\langle \delta h^2 \rangle$ are effective standard deviations of base pair twist and rise, between adjacent base pairs, away from their average values.

Thus, $\lambda_c^{(0)}$ is a measure of the level of distortion away from the ideal, undistorted, helix structure. For an ideal double helix, $\lambda_c^{(0)} = \infty$. In general, then, the difference in the interaction energies between homologous and non-homologous molecular pairs only becomes pronounced when the molecules are longer than $\lambda_c^{(0)}$. This is effectively the length over which the two non-homologous sequences can stay commensurate—strands on one molecule aligned with grooves on the opposite molecule—with each other (if not allowing for torsional flexibility), and maintain helix structure dependent interactions with each other. Hereafter, we use the value $\lambda_c^{(0)} \approx 150$ Å, which is within the range estimated in [22].

Note that $\phi_\mu^\alpha(z)$ also depends on thermal torsional fluctuations of the molecules. For short molecules this simply renormalizes $\lambda_c^{(0)}$ [10]. However, this can be neglected, as it is a relatively small correction. Another, less trivial effect is torsional adaptation where the helical structure of the molecules adjusts to facilitate better helix dependent interactions, due to the elasticity of the molecule [23, 24]. This effect on the pair interaction was studied in [16, 25–27]). Including these effects in molecular assemblies is mathematically and computationally involved for finite length molecules [28]. However, here we may, in the first approximation, neglect this effect by considering molecules that are not much longer than the thermal persistence length of DNA, like those used in the liquid-crystalline experiments [1].

Using equation (2.3), it is possible to recast equation (2.2) in a more practical form:

$$
E_{\text{int}} = \frac{1}{2} \int_{-L/2}^{L/2} \mathrm{d}z \sum_{\mu \neq \nu} \Big[a_0(|\mathbf{R}_\mu - \mathbf{R}_\nu|) - a_1(|\mathbf{R}_\mu - \mathbf{R}_\nu|) \cos\Big(\phi_\mu^{(0)} - \phi_\nu^{(0)} + \frac{s_\mu}{2}(1 - s_\mu s_\nu)\Delta\phi_\Omega(z) \Big)
$$
$$
+ a_2(|\mathbf{R}_\mu - \mathbf{R}_\nu|) \cos\Big(2\phi_\mu^{(0)} - 2\phi_\nu^{(0)} + s_\mu(1 - s_\mu s_\nu)\Delta\phi_\Omega(z) \Big), \tag{2.5}
$$

where

$$
\Delta\phi_\Omega(z) = \frac{1}{h} \int_0^z (\Omega^A(z') - \Omega^B(z')) \mathrm{d}z'. \tag{2.6}
$$

In equation (2.5), the freedom of choice between type *A* or *B* molecules has been characterised in terms of a spin-variable $s_\mu = \pm 1$; $s_\mu = 1$ corresponds to type *A* and $s_\mu = -1$ to type *B*.

Thus, the partition function (summing over both s_μ and ϕ_μ^0 degrees of freedom) can be defined as

$$
Z = \prod_{\mu=1}^{N} \sum_{s_\mu=-1,1} \int_0^{2\pi} \mathrm{d}\phi_\mu^0 \exp\Big(-\frac{E_{\text{int}}\{s_\mu, \phi_\mu^0\}}{k_B T} \Big). \tag{2.7}
$$

Potentially, for two specific base pair sequences, one should be able to determine the distortion patterns and calculate the partition function for those two sequences. However, our task is to calculate a more general, ensemble-averaged free energy, using the Gaussian statistics described previously. This average is defined as

$$
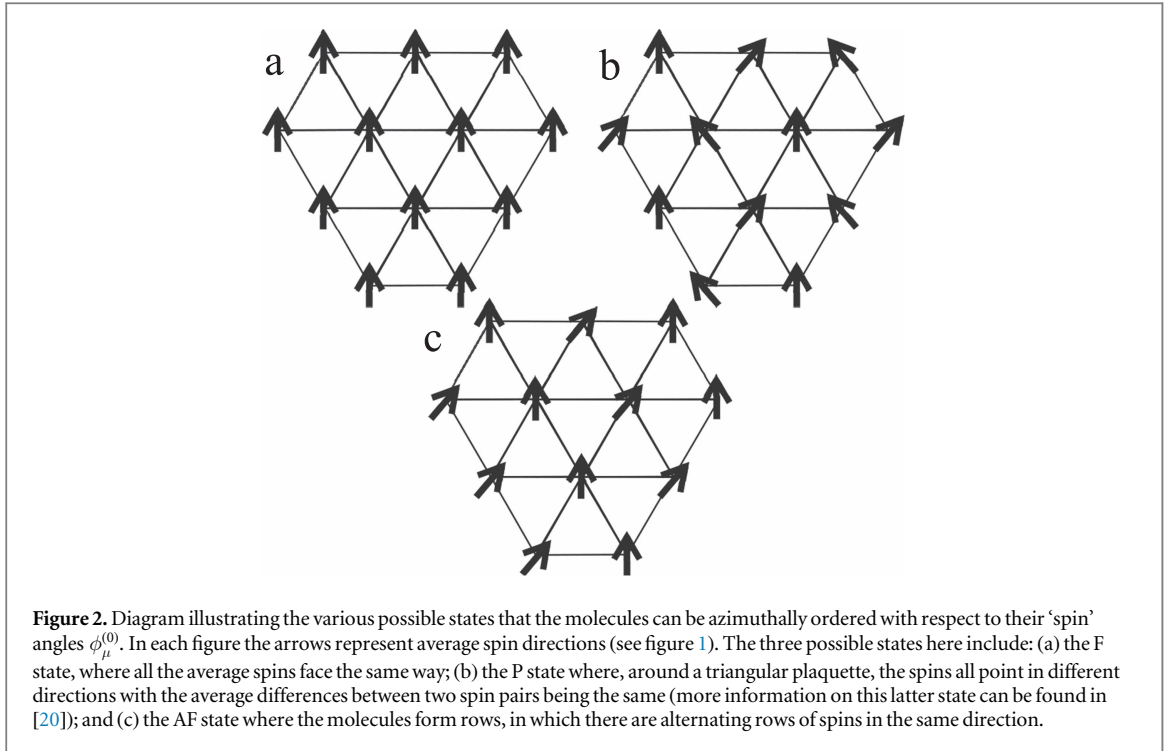\langle F \rangle_\Omega = \langle -k_B T \ln Z \rangle_\Omega. \tag{2.8}
$$

To calculate equation (2.8) precisely would be rather laborious and computationally expensive, involving detailed simulations that model specific patterns of helical distortions of many possible base-pair realisations, which then must be averaged. It might be possible to perform some kind of replica trick [29], on equation (2.8), to facilitate performing the ensemble average. However, this is not without its problems: firstly the ensemble average still cannot be performed analytically for this type of model, and secondly the analytic continuation in the number of replicas, *n* to zero is not a well-defined limit, and particular care must be taken. Instead, we utilise an approximation in which each pair potential for different helical distortion realizations is replaced with an ensemble-averaged one. This is effectively replacing quenched disorder with annealed disorder. This approximation automatically arises within the framework of a variational mean-field approximation, where we neglect correlations in s_μ. When the molecules are completely phase segregated or completely mixed those correlations are unlikely to be important; however, close to the transition such assumption may break down. The details of the variational approximation are fully presented in the supplemental material, whereas a more heuristic derivation of it, which relies on assuming annealed disorder, is presented in the main text. Thus, we write

$$
\langle F \rangle_\Omega \approx -k_B T \ln Z_{\text{eff}}, \tag{2.9}
$$

where

$$
Z_{\text{eff}} = \prod_{\mu=1}^{N} \sum_{s_\mu=-1,1} \int_0^{2\pi} \mathrm{d}\phi_\mu^{(0)} \exp\Big(-\frac{\langle E_{\text{int}}\{s_\mu, \phi_\mu^{(0)}\}\rangle_\Omega}{k_B T} \Big). \tag{2.10}
$$

The approximate free energy described by both equations (2.9) and (2.10) is also computed in MC simulations (for details see section 3 below). One may compute the leading order correction (and higher order corrections) to equation (2.9) in both simulations and the variational approximation, from considering the full average equation (2.8). Again, the method used to find these corrections is discussed in the supplemental material, which relies on an alternate to the replica trick. The average $\langle E_{\text{int}}\{s_\mu, \phi_\mu^{(0)}\}\rangle_\Omega$ is simple to obtain:

Figure 2. Diagram illustrating the various possible states that the molecules can be azimuthally ordered with respect to their 'spin' angles $\phi_\mu^{(0)}$. In each figure the arrows represent average spin directions (see figure 1). The three possible states here include: (a) the F state, where all the average spins face the same way; (b) the P state where, around a triangular plaquette, the spins all point in different directions with the average differences between two spin pairs being the same (more information on this latter state can be found in [20]); and (c) the AF state where the molecules form rows, in which there are alternating rows of spins in the same direction.

$$\langle E_{\text{int}}\rangle_\Omega = \frac{L}{2}\sum_{\mu\neq\nu}\left[a_0(|\mathbf{R}_\mu - \mathbf{R}_\nu|) - a_1(|\mathbf{R}_\mu - \mathbf{R}_\nu|)\left(1 - \frac{(1 - s_\mu s_\nu)}{2}\Psi\left(\frac{L}{2\lambda_c^{(0)}}\right)\right)\cos\left(\phi_\mu^{(0)} - \phi_\nu^{(0)}\right)\right.$$

$$\left. + a_2(|\mathbf{R}_\mu - \mathbf{R}_\nu|)\left(1 - \frac{(1 - s_\mu s_\nu)}{2}\Psi\left(\frac{2L}{\lambda_c^{(0)}}\right)\right)\cos\left(2\phi_\mu^{(0)} - 2\phi_\nu^{(0)}\right)\right], \tag{2.11}$$

where the function Ψ is defined as

$$\Psi(x) = 1 - \frac{1}{x}(1 - \exp(-x)). \tag{2.12}$$

The structure of equation (2.11), without the specific form for Ψ, is quite generic; thus its form may be also applicable to other candidate mechanisms for homologous pairing. To consider solely a local pairing mechanism (as described in introduction and also discussed in [6]) one could simply set $\Psi = 1$ in equation (2.11). Independently of this, as often is the case for statistical mechanical studies of specific systems, the overall model described by equations (2.9)–(2.11) may be of more general interest, as it describes a modified X–Y spin model describing particles with two types of magnetic interactions.

2.2. Possible phases and transitions in the model

The model potentially allows for many different states. First of all, we may have states of different azimuthal ordering of the molecules about their long axes. If the ratio $a_2(R)/a_1(R)$ is sufficiently small between nearest neighbours, we expect a state with $\langle\phi_\mu^{(0)} - \phi_\nu^{(0)}\rangle = 0$, where, on average, all the azimuthal angles ('spins') $\phi_\mu^{(0)}$ possess the same value. We call this state the 'ferromagnetic' (F) State. But, when the ratio $a_2(R)/a_1(R)$ becomes sufficiently large, states for which $\langle\phi_\mu^{(0)} - \phi_\nu^{(0)}\rangle = 0$ no longer holds everywhere may be energetically preferential. Two such examples are an 'antiferromagnetic' (AF) rhombic state and an antiferromagnetic 'Pott's' (P) State. In the simplest of these, the AF state, the molecules in the 2D lattice form rows or 1D layers. Along one particular lattice direction (within the row), nearest neighbour molecule's spins are parallel, $\langle\phi_\mu^{(0)} - \phi_\nu^{(0)}\rangle = 0$; however, the relative angle between neighbouring molecule's spins along other directions is non-zero, $\langle\phi_\mu^{(0)} - \phi_\nu^{(0)}\rangle = \psi$, where ψ takes a finite value (see figure 2). The P state has a slightly more complicated structure in terms of $\phi_\mu^{(0)}$, which is illustrated in figure 2. If we allow for rhombic lattice distortions, the AF state generally has a lower energy state compared to the P-state, for the values of parameters explored here. However, the free energies of these states are quite similar [20]. (If we fix the lattice as being triangular (hexagonal), the P-state will dominate at large values of $a_2(R)$, since the AF state is unstable without lattice distortions [20].)

Generally, we are interested in determining how phase segregation (separation of non-homologous molecules) depends on different molecular parameters. Segregation is driven by the size of $a_1(R)\Psi(L/2\lambda_c^{(0)})$ and

$a_2(R)\Psi(2L/\lambda_c^{(0)})$ for nearest neighbours in the lattice. When both $a_1(R)$ and $a_2(R)$ increase, so that the interaction energy between molecules dominates over mixing entropy, we might expect a phase almost pure in one type of molecule (complete segregation) as opposed to a mixed state containing significant numbers of both types. However, these two interaction terms may also compete against each other, and, therefore, lead to non-monotonic effects in the boundary between the mixed and segregated phases.

The degree of the phase segregation is also controlled by the function Ψ, which characterises the difference between the interaction between homologous and non-homologous molecules. Formally, we may expect that the largest degree of segregation occurs for very long molecules (as L goes to infinity). Here Ψ takes on its largest value, meaning the difference in the interaction between homologous pairs and non-homologous pairs is maximised. Note, however, that when the molecules are sufficiently large, the twisting elasticity (not included in the current model) would reduce this effect, allowing non-homologous molecules (again, molecules with different sequence text) to twist about their long axes in such a way to reduce their distortional mismatch, and hence their interaction energy [26]. On the other hand, for these long molecules, whose lengths are on the order of several bending persistence lengths, the liquid crystalline structures seen in the experiments of [1], generated to observe phase segregation, cannot be formed. And hence, we neglect this effect in our study as we desire to describe segregation effects only of (approximately) columnar liquid-crystalline mesophases of DNA.

2.3. Parameters for the KL model of DNA–DNA interactions

In our numerical calculations, the interaction coefficients, $a_0(R)$, $a_1(R)$, and $a_2(R)$, are determined from the KL model for helix structure specific DNA–DNA interactions [10, 11], which is based on a Debye–Bjerrum model (utilising mean-field electrostatics where ion-ion correlations are neglected). It entails the electrostatic interactions arising from the helical charge patterns on dsDNA. Here, the double helix is described by two phosphate backbones separated by two grooves—the smaller, minor groove and the larger, major groove (see figure 1). Each phosphate bears one negative charge, thus tracing out a double-helical charge distribution. Adsorbed counterions, cations, are assumed to be localised within these grooves, and therefore are considered as fixed charges on the molecular surface. These form helical positive charge charge distributions running along the grooves. All other ions in the system that are not bound to DNA provide linear Debye screening of these helical charge distributions. Within the framework of this model, the interaction terms $a_0(R)$, $a_1(R)$, and $a_2(R)$ are given by

$$a_0(R) = \frac{2l_{\mathrm{B}}(1-\theta)^2}{l_e^2}\frac{K_0(\kappa_{\mathrm{D}}R)}{(a\kappa_{\mathrm{D}}K_1(\kappa_{\mathrm{D}}a))^2} + \frac{2l_{\mathrm{B}}}{l_e^2}\sum_n[\xi_n(\theta, f_1, f_2)]^2\frac{\Omega_n(\kappa_n R, \kappa_n a)}{(a\kappa_n K_n'(\kappa_n a))^2}, \tag{2.13}$$

$$a_1(R) = \frac{4l_{\mathrm{B}}[\xi_1(\theta, f_1, f_2)]^2}{l_e^2}\frac{K_0(\kappa_1 R)}{(a\kappa_1 K_1'(\kappa_1 a))^2}, \tag{2.14}$$

$$a_2(R) = \frac{4l_{\mathrm{B}}[\xi_2(\theta, f_1, f_2)]^2}{l_e^2}\frac{K_0(\kappa_2 R)}{(a\kappa_2 K_2'(\kappa_2 a))^2}, \tag{2.15}$$

where

$$\Omega_n(\kappa_n R, \kappa_n a) = -\sum_j K_{n-j}(\kappa_n R)K_{n-j}(\kappa_n R)\frac{I_j'(\kappa_n a)}{K_j'(\kappa_n a)}, \tag{2.16}$$

and

$$\xi_n(\theta, f_1, f_2) = \theta f_1 + \theta f_2(-1)^n - \cos(n\tilde{\phi}_s), \quad \kappa_n = \sqrt{\kappa_{\mathrm{D}}^2 + n^2 g^2}. \tag{2.17}$$

Here, $I_n(x)$ and $K_n(x)$ are modified Bessel functions of the first and second kind, and $I'_n(x)$ and $K'_n(x)$ are their derivatives with respect to their arguments, respectively. The parameters of the model are: l_{B}, the Bjerrum length ($l_{\mathrm{B}} \approx 7$ Å); l_e, the mean distance between phosphate charges per unit charge ($l_e \approx 1.7$ Å); κ_{D}, the inverse Debye screening length (in all calculations we use the value $\kappa_{\mathrm{D}}^{-1} = 7$Å); $g = 2\pi/H$, where H is the average helical pitch of DNA ($H = 33.8$ Å); a is the effective electrostatic radius of DNA ($a \approx 11.2$ Å); $\tilde{\phi}_s$ is the angular half-width of the minor groove ($\tilde{\phi}_s \approx 0.4\pi$); θ is the degree of counterion condensation (or binding) near the DNA surface (if the charge of phosphates is 100% compensated with adsorbed counterions, $\theta = 1$); f_1 is the proportion of condensed (bound) ions localised in the minor groove; and f_2 is the proportion of condensed (bound) ions localised in the major groove.

We will also investigate the more generic phase behaviour of the phase diagram of our system, for a triangular lattice, in terms of a_1 and a_2, without their actual R-dependence being specified. Concentrating solely on these interaction terms realises results that are independent of the underlying interaction model, based on symmetry arguments presented above.

3. Methods

We calculate the free energy of these systems in order to determine when phase segregation occurs via two means: (i) a variational mean-field approximation and (ii) Monte Carlo simulations. We compare the results of the two methods to test their accuracy.

3.1. Variational approximation

In the variational approximation, when mixing between the two species occurs, we may use a 'jellium'-like, mean-field approximation. Here, we suppose that we can write the following mean-field ansatz (for the 'ferromagnetic' state) for the probability that a lattice site μ is occupied by either type A or B

$$P_\mu^\alpha \simeq (1 - \chi)\delta_{\alpha,A} + \chi\delta_{\alpha,B}. \tag{3.1}$$

Here, $\chi = \langle N_B\rangle/N$ is effectively an order parameter that characterises the degree of phase segregation, where N_B is number of lattice sites which are occupied by molecules of type B. In the model the total number of lattice sites N is fixed, and they are occupied either by A- or B-molecules, so that always $N_A + N_B = N$ is constant (N_A is number of lattice sites which are occupied by molecules of type A). Hence, we have $(1 - \chi) = \langle N_A\rangle/N$. Thus, it is evident that when $\chi = 0$ the system contains purely molecules of type A; whereas $\chi = 1$ corresponds to only type B; and when $\chi = 1/2$, there is 1:1 mixing of the two species.

Note that the above ansatz is shown to be equivalent to performing a more rigorous variational approximation using an effective 'magnetisation' (see supplemental material). However, using equation (3.1) is more physically transparent than this more sophisticated method, so we present it in the main text.

In the variational approximation, we may write an approximate free energy

$$F_T = -k_B T \ln Z_{T,0} + \langle\langle\langle E_{\text{int}}\rangle_\Omega\rangle_\alpha - E_T\rangle_T, \tag{3.2}$$

where

$$Z_{T,0} = \prod_{\mu=1}^N \int_{-\infty}^\infty \mathrm{d}\phi_\mu^{(0)} \exp\left(-\frac{E_T\{\phi_\mu^{(0)}\}}{k_B T}\right), \tag{3.3}$$

and

$$\langle\langle\langle E_{\text{int}}\rangle_\Omega\rangle_\alpha - E_T\rangle_T = \frac{1}{Z_{T,0}} \prod_{\mu=1}^N \int_{-\infty}^\infty \mathrm{d}\phi_\mu^{(0)} (\langle\langle E_{\text{int}}\{\phi_\mu^{(0)}\}\rangle_\Omega\rangle_\alpha - E_T\{\phi_\mu^{(0)}\}) \exp\left(-\frac{E_T\{\phi_\mu^{(0)}\}}{k_B T}\right). \tag{3.4}$$

The variational approximation is chosen to have a Gaussian (or harmonic) trial function where the finite range of values of ϕ_μ^0 is neglected, in both equations (3.3) and (3.4), and the limits of integration are extended to infinity. Here, the subscript α on one of the averaging brackets represents averaging over the effective probability given by equation (3.1). In the variational trial function, $E_T\{\phi_\mu^0\}$ are variational parameters that are chosen to minimise F_T to give the most accurate approximation to the full free energy (equation (2.9)).

For an illustration of the approximation method, we will consider in detail the F state. The trial energy function describing this state is given by

$$E_T = k_B T \sum_{\langle\mu,\nu\rangle} \frac{J}{2}(\phi_\mu^{(0)} - \phi_\nu^{(0)})^2, \tag{3.5}$$

where the angular brackets on the summation indices indicate that only nearest neighbour interactions have been considered. Indeed, J, an effective spring constant for relative rotations in $\phi_\mu^{(0)} - \phi_\nu^{(0)}$ for nearest neighbours, is chosen to minimise equation (3.2), which is the closest value to the exact free energy. The variational approximation can be extended to include next-to-nearest neighbours and so on.

The variational approximation yields, for the F state (considering only nearest neighbours),

$$\frac{F_T}{N} = 3a_0(R)L - 3a_1(R)L \exp\left(-\frac{1}{6J}\right)\left[1 - 2\chi(1-\chi)\Psi\left(\frac{L}{2\lambda_c^{(0)}}\right)\right]$$

$$+ 3a_2(R)L \exp\left(-\frac{2}{3J}\right)\left[1 - 2\chi(1-\chi)\Psi\left(\frac{2L}{\lambda_c^{(0)}}\right)\right] + \frac{k_B T}{2}\ln J + k_B T\hat{C}_{\text{hex}}. \tag{3.6}$$

When we use equation (3.1), in order to optimise χ we must also include a mixing entropy to realise a total free energy

$$F = F_T - TS_{\text{mix}}, \tag{3.7}$$

where

$$S_{\text{mix}} = -k_B N [(1 - \chi)\ln(1 - \chi) + \chi \ln(\chi)]. \tag{3.8}$$

The terms in F_T favour segregation of the two molecular types, while mixing entropy S_{mix} opposes this. The optimal χ balances these two effects.

In the more rigorous variational approximation, this additional entropy term is automatically included in the free energy and does not need to be added by hand (see supplemental material). Further details and how to formulate the variational approximation for the AF state (using a similar trial function to equation (3.5)) are given in the supplemental material. For the P state, we have not extended the mean-field approximation considered in [20] to consider mixing between the two molecule types. This is because the approximation for molecules of solely one type of base pair sequence (or ideal helices) is already very complicated. This is due to the reduced lattice symmetry of the P state. Additionally, it does not work very well due to thermal topological and domain-wall type excitations that were not incorporated [20]. Thus, MC simulations are required to probe the free energies for a fixed triangular lattice in which this P state occurs.

In appendix A of the supplemental material we present a simpler model, for ease of understanding, as it captures some key features of the analysis. Most importantly it describes the competition between mixing entropy and the interaction energy difference between homologous and non-homologous molecules. Similarities, as well as the disadvantages, of this model compared with the analysis used in the main text are discussed in the supplemental material.

3.2. MC Simulations

In the computational analysis, before applying the Monte Carlo Algorithm, the initial state of a system, at a given average nearest neighbour distance, was determined by performing a lattice sum of nearest-neighbour interactions. Here, the ground state energy was minimised upon varying (i) neighbouring spins of molecules, (ii) the type of pair interaction, being either of the homologous or non-homologous form , and (iii) the rhombic angle ω which provides the level of distortion of the rhombic lattice from a triangular one, where $\omega = 60°$.

The Monte Carlo simulations were carried out on a 24 $\times$ 24 site rhombic or triangular lattice with periodic boundary conditions for molecules of two different types. Here, only pair interactions (given by the terms in the summation of equation (2.11)) between nearest neighbours were considered. Three types of Monte Carlo move were considered: (i) altering the azimuthal angle $\phi_\mu^{(0)}$ of a molecule at a lattice site, (ii) changing the type (sequence) of a molecule (for instance, from type A to B) at a lattice site, or (iii) changing the rhombic angle ω of the lattice. Individual simulation runs were performed for systems of molecules with different lengths, densities, charges, and charge distributions. Furthermore, more generic simulations were carried out varying the interaction terms a_1 and a_2.
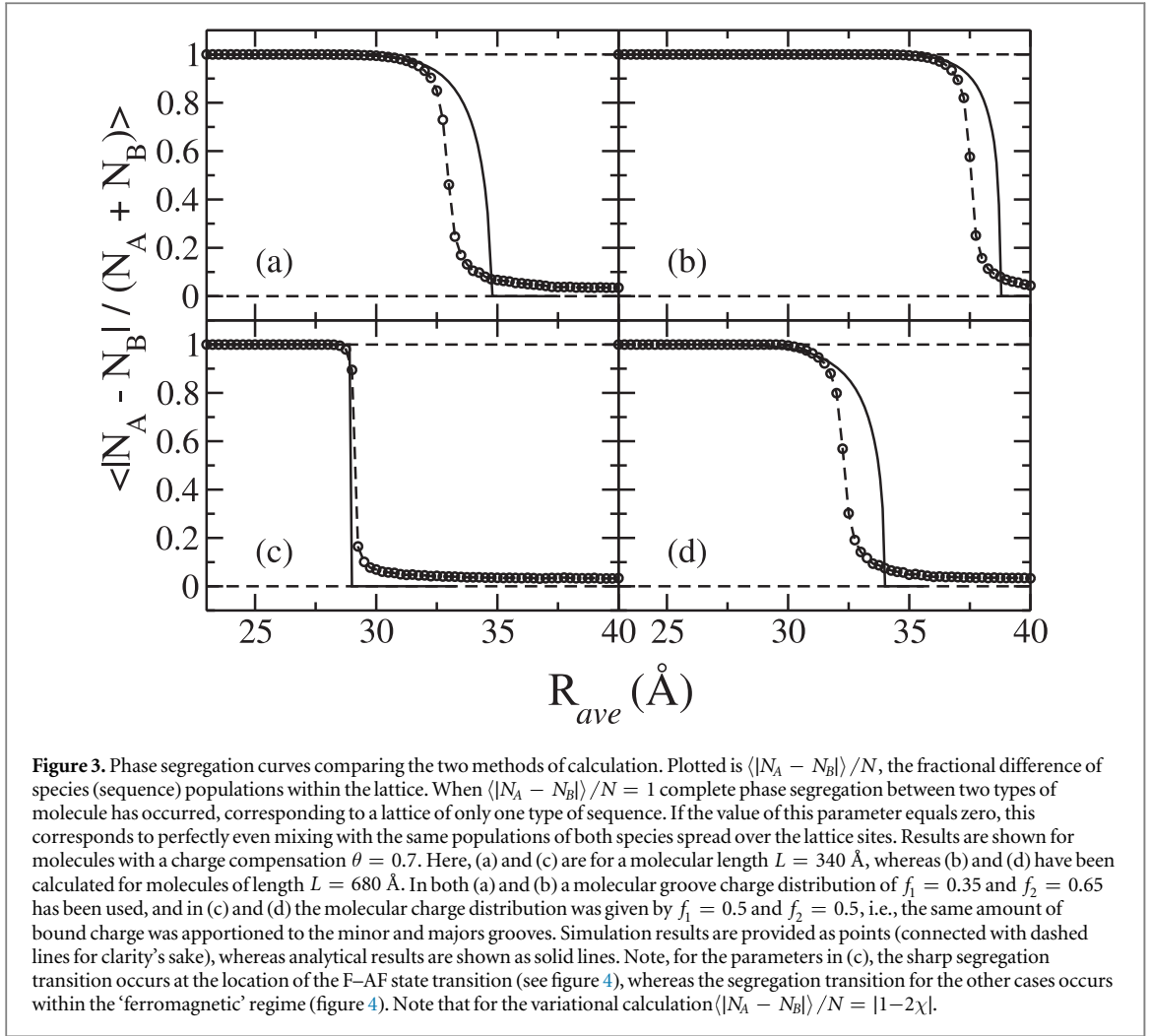
According to the Metropolis algorithm, MC moves were accepted with a probability of $\exp(-\Delta E/k_B T)$, if the energy of the new configuration was ΔE greater than that of the original configuration, and a probability of 1 if the energy of the new configuration was less than that of the old configuration. Systems were allowed to equilibrate for 10^5 MC steps before statistical averages of the states were calculated.

4. Results and discussion

4.1. Phase segregation and consistency between the two methods

In our results, we have determined the position of a mixing transition between a phase-segregated lattice state, with predominantly one type of molecule in the lattice, and a state in which two types of molecules are evenly mixed. This has been obtained for a range of interaction parameters for the KL model of DNA–DNA interactions. The transition line between these two states is defined as where $\langle |N_A - N_B| \rangle / N = 0.5$.

In determining the position of this mixing transition, both methods of calculation, the analytical variational approximation and the Monte Carlo simulations, agree well (see figure 3), indicating the reliability of these two approaches. However, each of the methods is not without its minor drawbacks. In the variational calculation, the system is treated at the mean-field level, thus we cannot expect it to be accurate at the transition point, since inhomogeneities and correlations in the spatial distribution of the two molecule types on the lattice are likely to become important here. On the other hand, the MC simulations were carried out on a relatively small lattice of molecules, and so finite-size effects may influence phase segregation behaviour in these simulations. For example, one clear artefact of a finite sized lattice, is that $\langle |N_A - N_B| \rangle / N$ remains significantly above zero for values of R_{ave}, the average interaxial distance between molecules, well above the transition. In fact, for the simulation results above the transition, the value of $\langle |N_A - N_B| \rangle / N$ remains constant as R_{ave} is increased. This constant value, arising from random ordering in a finite system, is calculated to be $\sqrt{2/\pi N}$, which, for the lattice size used in the simulations, is found to have a value of 0.033, consistent with what was found in the MC

Figure 3. Phase segregation curves comparing the two methods of calculation. Plotted is $\langle |N_A - N_B| \rangle /N$, the fractional difference of species (sequence) populations within the lattice. When $\langle |N_A - N_B| \rangle /N = 1$ complete phase segregation between two types of molecule has occurred, corresponding to a lattice of only one type of sequence. If the value of this parameter equals zero, this corresponds to perfectly even mixing with the same populations of both species spread over the lattice sites. Results are shown for molecules with a charge compensation $\theta = 0.7$. Here, (a) and (c) are for a molecular length $L = 340$ Å, whereas (b) and (d) have been calculated for molecules of length $L = 680$ Å. In both (a) and (b) a molecular groove charge distribution of $f_1 = 0.35$ and $f_2 = 0.65$ has been used, and in (c) and (d) the molecular charge distribution was given by $f_1 = 0.5$ and $f_2 = 0.5$, i.e., the same amount of bound charge was apportioned to the minor and majors grooves. Simulation results are provided as points (connected with dashed lines for clarity's sake), whereas analytical results are shown as solid lines. Note, for the parameters in (c), the sharp segregation transition occurs at the location of the F–AF state transition (see figure 4), whereas the segregation transition for the other cases occurs within the 'ferromagnetic' regime (figure 4). Note that for the variational calculation $\langle |N_A - N_B| \rangle /N = |1 - 2\chi|$.
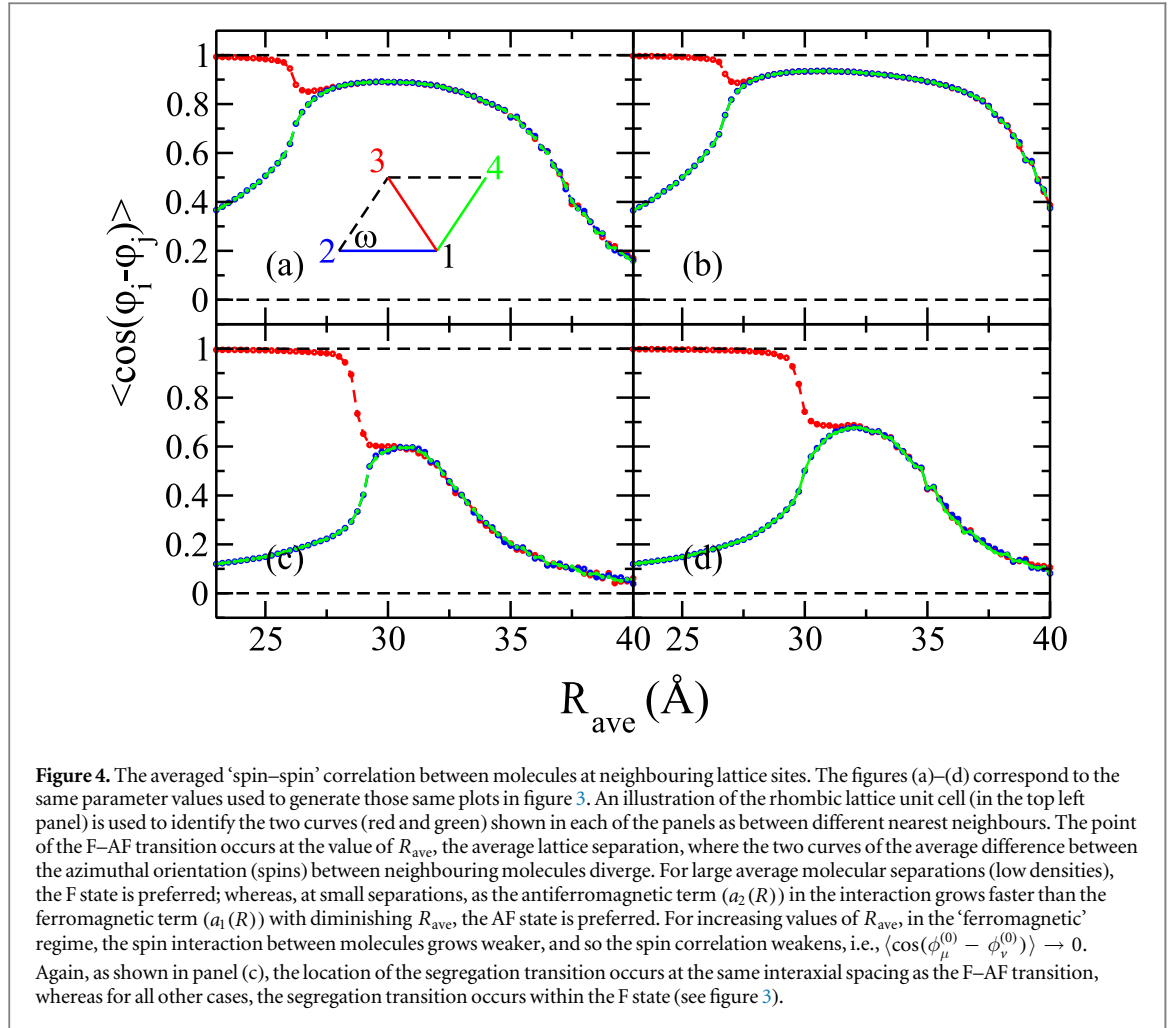
simulations. As shown in figure 3, near the transition, the curves for the simulations in figure 3 are much less steep than those of the variational calculation. But it is well known that finite size effects generally smear transitions, which is precisely what is observed in the finite-sized MC simulations. We presume that exact results for $\langle |N_A - N_B| \rangle /N$ lie somewhere between these two curves.

This mixing transition, as shown, depends on molecule length and the distribution of charge on the molecular surface as provided in the interaction model. The effect of increasing length is to push the transition to larger values of R_{ave}, which thus favours the segregated state over a larger density range in these DNA assemblies. Indeed, this result is not surprising. According to our model, the difference between the average interaction energy between homologous and non-homologous DNA increases as the molecular length increases, and hence, this increasing difference would increase the density range over which phase segregation occurs. Of the two ion distributions shown in figure 3, we find the distributions with the larger value of f_1 (a greater fraction of ions localised in the minor groove) yield transitions at smaller interaxial distances R_{ave}. As shown in figure 5, this is a general trend, which we will discuss below.

4.2. The AF–F transition

Besides the transition between the mixed and segregated phases of molecules with different sequences, there is also an 'antiferromagnetic'–'ferromagnetic' (AF–F) transition, as discussed in section 2.2. This second transition changes the ordering of the azimuthal orientations of the centres of the DNA helices about their long axes. At large values of R_{ave} (when $\langle \phi_\mu^{(0)} - \phi_\nu^{(0)} \rangle = 0$ for all nearest neighbours), the lowest energy state corresponds to the F state. As R_{ave} decreases, a_2 (in equation (2.11)) that favours $\langle \phi_\mu^{(0)} - \phi_\nu^{(0)} \rangle \neq 0$ grows faster relative to a_1, which then results in the AF rhombic lattice (or P state for molecules fixed to a triangular lattice) being the lowest energy state. Again, for the AF state, the molecules are azimuthally ordered in rows. The average value of the azimuthal angle difference between neighbouring molecules $\langle \phi_\mu^{(0)} - \phi_\mu^{(0)} \rangle = 0$ within a single row of molecules, while, along other lattice directions, between nearest neighbours (i.e. between rows), $\langle \phi_\mu^{(0)} - \phi_\nu^{(0)} \rangle = \psi$, where ψ is some no-zero value. (Again, for details, see section 2.2.)

Figure 4. The averaged 'spin–spin' correlation between molecules at neighbouring lattice sites. The figures (a)–(d) correspond to the same parameter values used to generate those same plots in figure 3. An illustration of the rhombic lattice unit cell (in the top left panel) is used to identify the two curves (red and green) shown in each of the panels as between different nearest neighbours. The point of the F–AF transition occurs at the value of R_{ave}, the average lattice separation, where the two curves of the average difference between the azimuthal orientation (spins) between neighbouring molecules diverge. For large average molecular separations (low densities), the F state is preferred; whereas, at small separations, as the antiferromagnetic term ($a_2(R)$) in the interaction grows faster than the ferromagnetic term ($a_1(R)$) with diminishing R_{ave}, the AF state is preferred. For increasing values of R_{ave}, in the 'ferromagnetic' regime, the spin interaction between molecules grows weaker, and so the spin correlation weakens, i.e., $\langle \cos(\phi_\mu^{(0)} - \phi_\nu^{(0)}) \rangle \to 0$. Again, as shown in panel (c), the location of the segregation transition occurs at the same interaxial spacing as the F–AF transition, whereas for all other cases, the segregation transition occurs within the F state (see figure 3).

To probe this transition, as well as to measure the amount of azimuthal fluctuations in the lattice, we consider the 'spin–spin' correlation function $\langle \cos(\phi_\mu^{(0)} - \phi_\nu^{(0)}) \rangle$ between nearest neighbours. As fluctuations increase, the value of $\langle \cos(\phi_\mu^{(0)} - \phi_\nu^{(0)}) \rangle$ decreases. Indeed, for large interaxial spacings between molecules, we expect both the size of a_1 and a_2 in equation (2.11) to exponentially decrease (see equations (2.14) and (2.15)) with increasing R_{ave}, leading to a weaker interaction between the molecules and, hence, a decrease in the 'spin'–'spin' correlation function as shown in figure 4. for large R_{ave}.
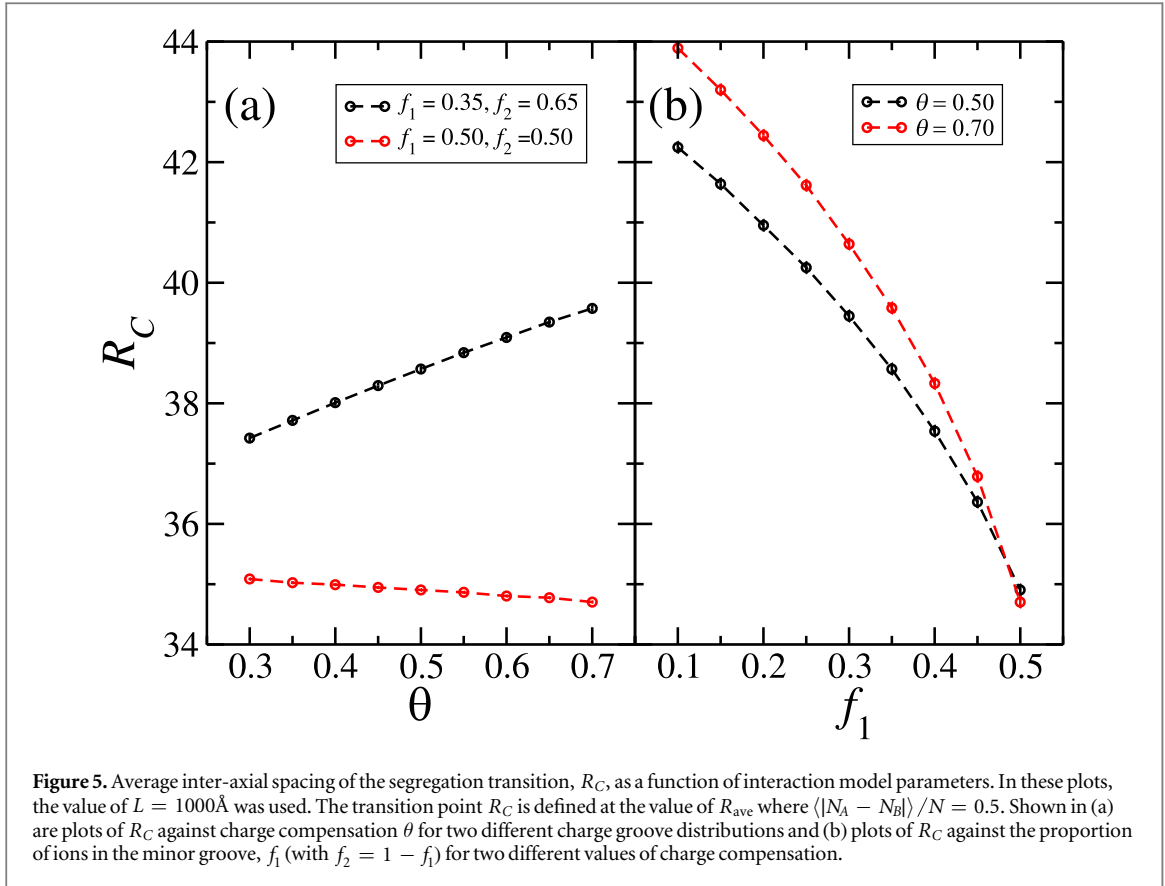
In the AF state, between the rows of molecules with different 'spins' (see figure 2), we can write

$$\langle \cos(\phi_\mu^{(0)} - \phi_\nu^{(0)}) \rangle = \cos\psi \langle \cos(\phi_\mu'^{(0)} - \phi_\nu'^{(0)}) \rangle. \tag{4.1}$$

Here, $\phi_\mu^{(0)} - \phi_\nu^{(0)} = \psi + \phi_\mu'^{(0)} - \phi_\nu'^{(0)}$ and the values $\phi_\mu'^{(0)} - \phi_\nu'^{(0)}$ fluctuate about the value zero, i.e. $\langle \phi_\mu'^{(0)} - \phi_\nu'^{(0)} \rangle = 0$. Indeed $\langle \cos(\phi_\mu'^{(0)} - \phi_\nu'^{(0)}) \rangle$ is reduced by thermal fluctuations. In equation (4.1), as ψ increases, $\langle \cos(\phi_\mu^{(0)} - \phi_\nu^{(0)}) \rangle$ must also decrease. Thus, in the AF state, between rows with different 'spins' we expect that $\langle \cos(\phi_\mu^{(0)} - \phi_\nu^{(0)}) \rangle$ is smaller than its value along the rows. Due to this difference, in figure 4, the F–AF transition takes place where the red and green curves of $\langle \cos(\phi_\mu^{(0)} - \phi_\nu^{(0)}) \rangle$, for nearest neighbours in the two lattice directions, as functions of R_{ave} diverge.

By comparing both figures 3 and 4 we may judge the relative positions of the mixing transition with respect to the F–AF transition. For the parameter values $f_1 = 0.35$ and $f_2 = 0.65$, the mixing transition occurs well within the F state of the molecules, whereas, for $f_1 = 0.5$ and $f_2 = 0.5$ (an even distribution of bound counterions distributed in the minor and major grooves), the mixing transition still occurs well within the F state for longer molecules ($L = 680$ Å), but for the smaller molecules ($L = 340$ Å), the mixing transition occurs at the same interaxial molecular spacing as the AF–F transition.

Increasing the ratio f_1/f_2, i.e., the fraction of bound ions in the minor groove to that in the major groove, has the effect of pushing the AF–F transition to larger interaxial spacing, R_{ave} since doing so increases the size of the ratio $a_2(R)/a_1(R)$ that controls the transition, as it is the relative proportion of the two interaction terms that dictates the strength of the antiferromagnetic aspect of the interaction. Furthermore, increasing the ratio of

Figure 5. Average inter-axial spacing of the segregation transition, R_C, as a function of interaction model parameters. In these plots, the value of $L = 1000$Å was used. The transition point R_C is defined at the value of R_{ave} where $\langle |N_A - N_B| \rangle / N = 0.5$. Shown in (a) are plots of R_C against charge compensation θ for two different charge groove distributions and (b) plots of R_C against the proportion of ions in the minor groove, f_1 (with $f_2 = 1 - f_1$) for two different values of charge compensation.

$a_2(R)$ to $a_1(R)$ essentially reduces the strength of the interaction (in the F state) since these terms come into the interaction with opposite signs. Therefore, the mixing transition moves down to smaller interaxial spacing as the $a_2(R)/a_1(R)$ ratio grows, since here, thermal fluctuations grow, leading to increased mixing of different molecular species. The position of the F–AF transition is insensitive to L. However, increasing the length, L, of the molecules does significantly reduce the degree of $\phi_\mu^{(0)}$ fluctuations, as can be seen in figure 4.
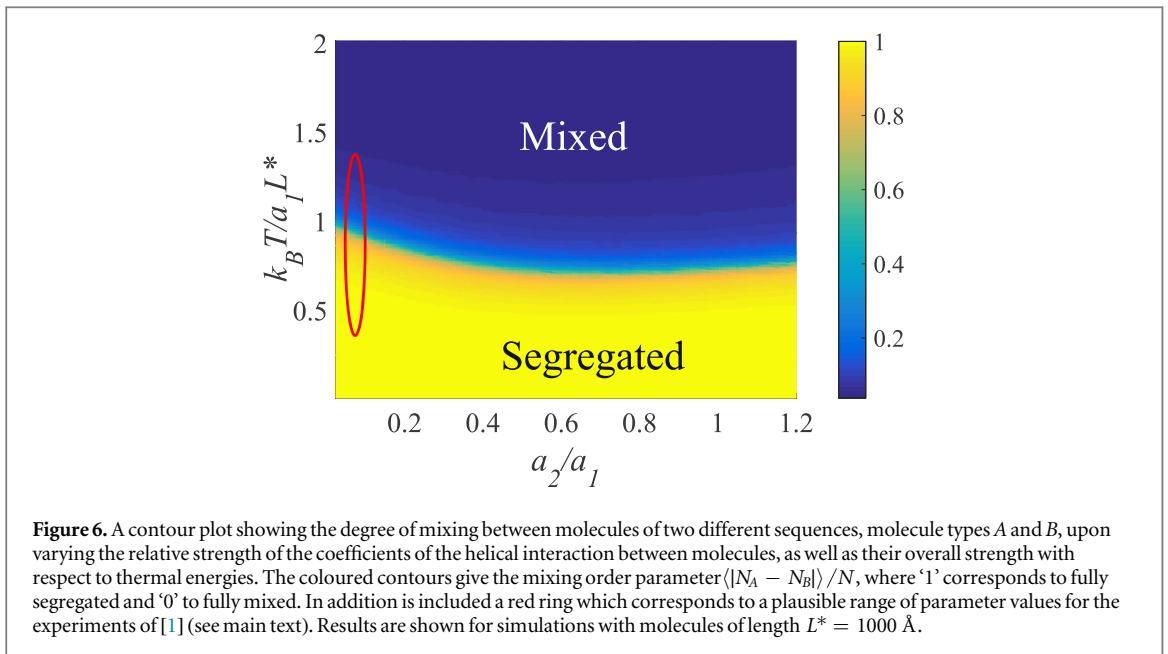
The plots shown in figure 4 are found from the MC simulations, where we observe a smooth transition between the segregated and mixed states. The variational calculation yields a discontinuous AF-F transition (see [20]), which more or less occurs around where the curves in figure 4 show a more dramatic weakening in the spin correlation between the molecules. The reasons for the main discrepancies result, again, from the finite size effects of the simulations or the mean-field elements of the variational theory breaking down very close to the point of transition.

4.3. The dependence of segregation on KL-model parameters

For further exploration of the effects of changing the counter-ion distribution close to the surface of the molecules, we specifically analyse results for molecules of a particular (average) length, $L = L^*$, where $L^* = 1000$ Å or 294 base pairs. This was the length of molecules used in the segregation experiments of [1]. For these length molecules, we then find the location of the mixing transition location, defined as $R_C = R_{ave}$ where $\langle |N_A - N_B| \rangle / N = 0.5$, upon varying θ (the counterion charge compensation of the DNA molecule) and f_1 and f_2 (again, the values that denote how bound or adsorbed counterions are apportioned between the minor and major grooves). These terms were varied because the effective DNA charge and distribution of counter-ions were found to depend on the species and concentrations of ions used in experiments. In these results we suppose that $f_1 + f_2 = 1$, which should be more or less the case for ions that bind strongly to the grooves.

To describe the experiments of [1], $f_1 + f_2$ is not precisely known, however, but it is likely smaller than 1 considering the monovalent salt solutions used in these experiments. However, a DNA charge compensation value of $\theta = 0.3-0.4$ could be predicted for the experiments upon fitting the osmotic pressure data of DNA osmotic pressure experiments under similar conditions [30, 31]. Finally, these fits revealed that the results were rather insensitive to the value of $f_1 + f_2$ for these smaller charge compensations θ. Later we consider a set of, more or less, realistic values of model parameters to describe those experiments.

In figure 5(a), results are shown for systems in which the DNA charge compensation, θ, is varied for two sets of charge groove distributions, f_1 and f_2 values, namely $f_1 = 0.5$, $f_2 = 0.5$ and $f_1 = 0.35$, $f_2 = 0.65$. In the case

Figure 6. A contour plot showing the degree of mixing between molecules of two different sequences, molecule types *A* and *B*, upon varying the relative strength of the coefficients of the helical interaction between molecules, as well as their overall strength with respect to thermal energies. The coloured contours give the mixing order parameter $\langle |N_A - N_B| \rangle / N$, where '1' corresponds to fully segregated and '0' to fully mixed. In addition is included a red ring which corresponds to a plausible range of parameter values for the experiments of [1] (see main text). Results are shown for simulations with molecules of length $L^* = 1000$ Å.

where $f_1 = 0.35$, $f_2 = 0.65$, the interaxial spacing of the mixing transition, R_C, increases strongly with θ. The helical electrostatic coefficient, a_1, of the interaction increases as more counter-ions are localised to the helical grooves (θ becoming larger). Therefore, at a given average separation, R_{ave}, this increase results in stronger preferential interactions between identical molecules, which dominates the thermal fluctuations at separation values below the mixing transition. It is this that pushes the mixing transition to a larger interaxial spacing R_C. However, when the counterion bound charge is more evenly distributed between the grooves, as seen for the values $f_1 = 0.5$, $f_2 = 0.5$, the effects of varying θ are weakened by the increase in the ratio $a_2(R)/a_1(R)$. Again, since $a_2(R)$ comes in with an opposite sign with respect to $a_1(R)$ in the interaction, the larger value of $a_2(R)$ effectively weakens the interaction, and, hence, the impetus for segregation, so that the system remains mixed until the assemblies are at larger densities (smaller values of R_{ave}), with the increase of θ.

We can also fix θ and vary f_1, holding $f_2 = 1 - f_1$, such that counterions are bound either to the minor or major grooves. In figure 5(b) results are shown for two fixed values of the charge compensation, $\theta = 0.5$ and $\theta = 0.7$. In both cases, we find that value of R_C depends dramatically on the value of f_1. Again, as more counter-ion charge is localised in the minor groove (f_1 is increased) the ratio of $a_2(R)/a_1(R)$ increases, effectively reducing the difference in interaction energies between homologous and non-homologous pairs. This then hinders phase segregation, such that, again, R_C, the location where segregation occurs, occurs at smaller intermolecular spacing. In figure 5(b), for the larger value, $\theta = 0.7$, we find that the mixing transition occurs at greater interaxial separations than for the smaller charge compensation, $\theta = 0.5$, for most values of f_1 considered, due to a reduction in overall interaction strength. However, this difference becomes smaller as f_1 is increased, and reverses at values above $f_1 \approx 0.5$. This is again governed by the increasing ratio of $a_2(R)/a_1(R)$, as was explained above for figure 5(a).

4.4. The $a_1 - a_2$ phase diagram for a fixed triangular lattice
As well as simply focusing on the KL model of interaction, we also show results that are independent of this model, though relying on the global mechanism of homology recognition through helix distortions. Here we show the degree of phase segregation as a function of the parameter values a_1 and a_2 for helix structure specific interactions. Results are shown, again, for molecules with the length used in the experimental segregation studies [1].

This figure illustrates some features discussed in the previous sections. Firstly, as the helix-structure-dependent interaction increases, (i.e., a_1 increases), which occurs with increased aggregate density or charge localisation within the DNA grooves, the relative difference in the interaction energies between homologous and non-homologous increases, favouring segregation. Secondly, as more counter-ion charge is apportioned to the minor groove, increasing the relative value of the antiferromagnetic a_2 term to the ferromagnetic a_1 term, the effective helical interaction decreases. Hence, the difference of the interaction energies, again, between homologous and non-homologous molecular pairs decreases, suppressing phase segregation so that it occurs at even larger values a_1 (and thus, at larger aggregate densities).

Interestingly enough, as the ratio of a_2/a_1 is further increased, this trend reverses slightly (see the behaviour of the 'segregation line' in right side of the figure 6). Here, the effective helical interaction increases again and,

hence, the difference of the interaction energies between homologous and non-homologous molecules. This effect may be linked to the fact that the ground state here is now the P state. This azimuthal ordering is obscured by fluctuations, however it still may contribute to free energy, although on average it is not the most popular state: the mixing transition, for this molecular length, always lies on average in the F state. For this reason this opposite trend in figure 6 is very weak. Indeed, we would expect for P state that increasing a_2 would increase the effective strength of interactions.

We find that phase transitions between the topologically disordered and ordered spin P-states [20] lie in the phase segregated region for molecules of chosen length L^*. The phase diagram for such spin transitions is described in [20]. The effect of reducing the length would be to bring this spin transition closer to the phase segregation line and eventually, crossing it, which might lead to interesting effects, as observed for the shorter molecules in figures 3 and 4. On the other hand, phase segregation for $L = L^*$ was already weak under the conditions explored in [1].

In the phase segregation transition plot, figure 6, we also show the approximate region in the $a_1 - a_2$ parameter space that may correspond to the segregation experiments of [1], using the K-L model of DNA-DNA interaction. To determine the possible range of values of these parameters, fits were performed on results of osmotic pressure experiments of DNA aggregates [30, 31] under similar experimental conditions as those found in [1]. With these fits, estimates of $\theta \approx 0.3-0.4$ could be deduced for the level of charge compensation of DNA by adsorbed counterions. Appropriate values for f_1 and f_2, the fraction of counterions adsorbed in the minor and major grooves, respectively, could not as easily be deduced from the osmotic pressure experiments. However, with the relatively small value of the estimated counterion charge compensation found from these fits, the influence on the interaction terms a_1 and a_2 of the distribution of counterion adsorption, provided by f_1 and f_2, is rather small. Thus, to limit the number of parameters, we assumed a random counterion distribution along the DNA molecule, which corresponds to values of f_1 and f_2 being 0.

A larger influence is exerted on the interaction terms by the separation between neighbouring molecules, since these terms decrease exponentially with molecular separation. The molecular separations had not been measured in those experiments. However, since the aggregates in [1] were found to be in a mild cholesteric state, we could surmise that the average interaxial separation between neighbouring molecules was roughly 3–4 nm, which corresponds roughly to a surface-to-surface separation of 1–2 nm.

With the estimates for charge compensation θ and average molecular interaxial separation (along with the values of $f_1 = f_2 = 0$), using equations (2.14) and (2.15), we can determine the possible range of values for the interaction terms, a_1 and a_2 that correspond to the conditions of the segregation experiments [1]. Because of, mainly, the large uncertainty in the distances between neighbouring molecules, there, in turn, was a large uncertainty in these values, with $a_1 L^*$ varying anywhere from $0.4-1.4k_B T$, and a_2 varying from $0.05-0.2a_1$. This experimental region, drawn in figure 6, straddles the segregation transition boundary of the simulations. It encompasses the mixing ratio range $\langle |N_A - N_B| \rangle /N \approx 0.1-0.5$ found, by quantifying the degree of red-green colour separation, in the experiments. Of course, on the simulation side, there is still a degree of uncertainty in the precise location as well as the width of the mixing transition since, again, finite size effects of MC simulations have the tendency to broaden the transition. And, of course, on the experimental side, we have not accounted for smearing of the transition in experiments due to non-equilibrium effects during the formation of these DNA mesophases.

5. Concluding remarks

In this article we have used variational approximations and computer simulations to establish the nature of segregation of DNA molecular fragments in a model for phase segregation of two DNA types. We have shown how this segregation depends on a variety of molecular parameters that characterise the difference in the interaction energies between fragments of homologous DNA molecules, sequences with the same base pairs, and non-homologous molecular pairs, which have different, totally uncorrelated sequences. Here, to follow one specific line of enquiry, we have assumed that the difference of energies relies exclusively on the base-pair sequence-dependent patterns of helical distortions away from an ideal helix. This is not to say that there could be mechanisms relying on local base pair dependent interactions [6, 17, 18] playing a role. The extent of such a role has yet to be ascertained. However, it should be relatively easy to adapt this general statistical model to take account of such mechanisms, once analytical models of interaction have been proposed.

As well as phase segregation, our model also predicts different states in which the centres of the molecules can be azimuthally ordered, which can be described in the statistical mechanical language of magnetic systems (the helical interaction between DNA pairs depends on the relative azimuthal orientation (or 'spin') of the molecules about their long axes). Indeed, it is quite possible that our model may also have application in the study of magnetic systems, when two types of magnetic species are considered. Of particular interest to us was

the relative molecular separation at which the mixing transition between DNA molecules (between the two types of sequences) occurs with respect to that of magnetic-like transitions of the molecular spin states. Granted, detection of these 'spin-ordered' DNA states may be difficult to detect experimentally, unless synchrotron x-ray diffraction experiments are invoked. On the other hand, such changes in the azimuthal ordering could affect the phase segregation of the two molecular types under certain conditions. Generally, though, for realistic model parameters, we found that this mixing transition occurs when the DNA molecules azimuthal orientations are aligned on average, i.e., they are in language of magnetic spins in a 'ferromagnetic state'. However, we discovered non-trivial effects of the helical interaction on the location of the mixing transition. Namely, the a_2 helical harmonic that favours the antiferromagnetic state in our helical interaction model may reduce the effective strength of helical interactions. This reduces the difference of the interaction energies between homologous and non-homologous molecular pairs (the 'molecular recognition energy'), which, in turn, increases the likelihood of mixing of molecules of different sequences. Also, we found that if that if the 'anti-ferromagnetic' term, a_2 could be made strong enough this effect could be weakly reversed.

We have compared the predictions of this model, using the KL theory of interaction, with the observations of the experimental study that first reported segregation [1]. Using parameter values that fit the osmotic pressure experiments of [30, 31] and values for the intermolecular densities of these DNA assemblies, as suggested by the formation of a mild cholesteric state in experiments of [1], we roughly determined the corresponding interaction parameters of the experiments so that the experimental region could be mapped onto the predicted phase diagram for mixing/segregation. This mapping revealed a large predicted range for the mixing order parameter $\langle |N_A - N_B| \rangle / N$, as predicted by the theory, which completely encompasses the range of these values, $\langle |N_A - N_B| \rangle / N \approx 0.1-0.5$, found in the experiments.

This indicates that the model is indeed plausible for these weakly cholesteric condensed DNA spherulite systems, although the large degree in the uncertainty in determining the interaction terms for the experiments limits comparisons to theory. Unfortunately, again, in those experiments, the exact average value of lattice spacing R for the spherulites was not measured using x-ray diffraction or some other appropriate method, nor was the osmotic stress varied with changing PEG concentration. Thus, we cannot better fit our model to experiments, or vice versa. Therefore, new experiments are clearly needed to determine the dependence of the degree of phase segregation on R in the spherulites.

We envision that other relatively simple experiments could also be undertaken to study how varying other molecular parameters of DNA would influence segregation. For example, we would expect that increasing the length of the molecules would also increase the degree of segregation. Thus, different lengths of DNA fragments that form a liquid crystal phase should be investigated. Furthermore, we might also expect that changing the background electrolyte (type or concentration) in these experiments, which would alter the charge environment of the DNA, could also impact segregation by changing the distribution of adsorbed (or bound) ions about the DNA . Lastly, one might investigate to what degree changing the base-pair sequence would affect the phase segregation, and whether this correlates with expected patterns of helical distortions.

These proposed systematic experimental investigations should allow us to better understand the interactions involved in homology recognition. Indeed, our first goal would be to try to fit the parameters of this current model to such data, before considering other possible candidate mechanisms of homology recognition [6, 17, 18]. Notably, such studies would be far-reaching, as the nature of such forces may be important in initiating critical biological processes such as homologous recombination.

Acknowledgments

References

[1] Baldwin G, Brooks N, Robson R, Goldar A, Wynveen A, Leikin S, Seddon J and Kornyshev A A 2008 Duplex DNA recognize sequence homology in protein free environment *J. Phys. Chem.* B **112** 1060
[2] 2008 Spooky attraction of DNA from a distance *New Sci.* 15
    2008 Paired pairs *Nature* **451** 609
    2008 DNA's self regard *Science* **319** 879
    2008 Seeking recognition *Biopolymers* **89** 4
    Choi C Q 2008 Double helix double up *Sci. Am.* **298** 30

Falaschi A 2008 Commentary: Similia similibus: pairing of homologous chromosomes driven by the physicochemical properties of DNA *HFSP J.* **2** 257

[3] Inoue S, Sugiyama S, Travers A A and Ohyama T 2007 Self-assembly of double stranded DNA molecules at nanomolar concentrations *Biochemistry* B **46** 164

[4] Danilovicz C, Lee C H, Kim K, Hatch K, Coljee V W, Kleckner N and Prentiss M 2009 Single molecule detection of direct, homologous DNA/DNA pairing *Proc. Natl Acad. Sci. USA* **106** 19824

[5] Nishikawa J and Ohyama T 2013 Selective association between nucleosome with identical DNA sequences *Nucleic Acids Res.* **41** 1544

[6] Lee D J (O'), Danilowicz C, Rochester C, Kornyshev A A and Prentiss M 2016 Evidence of protein-free homology recognition in magnetic bead force–extension experiments *Proc. R. Soc.* A **472** 20160186

[7] Cha R S, Weiner B M, Keeney S, Dekker J and Kleckner N 2000 Progression of meiotic DNA replication is regulated by interchromosomal interaction proteins, negatively by Spo11p and positively by Rec8p *Genes Dev.* **14** 493

[8] Zickler D 2006 From early homologue recognition to synaptonemal complex formation *Chromosoma* **115** 158

[9] Barzel A and Kupiec M 2008 Finding a match: How do homologous sequences get together for recombination? *Nat. Rev. Genet.* **9** 27

[10] Kornyshev A A, Lee D J, Leikin S and Wynveen A 2007 Structure and interactions of biological helices *Rev. Mod. Phys.* **79** 943

[11] Kornyshev A A and Leikin S 1999 Electrostatic zipper motif for DNA aggregation *Phys. Rev. Lett.* **82** 4138

[12] Dickerson R E and Drew H R J 1981 Structure of a B-DNA dodecamer: II influence of base sequence on helix structure *J. Mol. Biol.* **149** 761

[13] Dickerson R E 1992 DNA structure from A to Z *Methods Enzymol.* **211** 67

[14] Kornyshev A and Leikin S 2001 Sequence recognition in the pairing of DNA duplexes *Phys. Rev. Lett.* **86** 366

[15] Kornyshev A A 2010 Physics of DNA: unravelling hidden abilities encoded in the structure of 'the most important molecule' *Phys. Chem. Chem. Phys.* **39** 12352

[16] Lee D J, Wynveen A, Albrecht T and Kornyshev A A 2015 Which way up? Recognition of homologous DNA segments in parallel and antiparallel alignment *J. Chem. Phys.* **142** 045101

[17] Mazur A K 2016 Homologous pairing between long DNA double helices *Phys. Rev. Lett.* **116** 158101

[18] Lu B, Naji A and Podgornik R 2015 Molecular recognition by van der Waals interaction between polymers with sequence-specific polarizabilities *J. Chem. Phys.* **142** 214904

[19] Lavery R R, Maddocks J H, Pasi M and Zakrzewska K 2014 Analyzing ion distributions around DNA *Nucleic Acids Res.* **42** 8138

[20] Wynveen A, Lee D J and Kornyshev A A 2005 Statistical mechanics of columnar DNA assemblies *Eur. Phys. J.* E **16** 303

[21] Kornyshev A A and Wynveen A 2009 The homology recognition well as an innate property of DNA structure *Proc. Natl Acad. Sci. USA* **106** 4683

[22] Wynveen A, Lee D J, Kornyshev A A and Leikin S 2008 Helical coherence of DNA in crystals and solution *Nucleic Acids Res.* **36** 5540

[23] Lipfert J, Kerssemakers J W J, Jager T and Dekker N H 2010 Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments *Nat. Methods* **7** 977

[24] Smith S B, Cui Y J and Bustamante C 1996 Overstretching B-DNA: the elastic response of individual double-stranded and single-strand DNA molecules *Science* **271** 5250

[25] Kornyshev A A and Wynveen A 2004 Nonlinear effects in torsional adjustment of interacting DNA *Phys. Rev.* E **69** 041905

[26] Cherstvy A G, Kornyshev A A and Leikin S 2004 Torsional deformation of double helix in interaction and aggregation of DNA *J. Phys. Chem.* B **108** 6508

[27] Lee D J, Wynveen A and Kornyshev A A 2004 DNA–DNA interaction beyond the ground state *Phys. Rev.* E **70** 051913

[28] Lee D J and Wyneveen A 2006 Torsional fluctuations in columnar DNA assembles *J. Phys.: Condens. Matter* **18** 787

[29] Mezard M, Parisi G and Virosoro M 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)

[30] Rau D C and Paresgian V A 1992 Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces *Biophys. J.* **61** 246

[31] Rau D C and Paresgian V A 1992 Direct measurement of temperature-dependent solvation forces between DNA double helices *Biophys. J.* **61** 260