Soft Matter

Cite this: *Soft Matter*, 2012, **8**, 8388

www.rsc.org/softmatter

PAPER

Self-assembly of short DNA duplexes: from a coarse-grained model to experiments through a theoretical link

Cristiano De Michele,*[a] Lorenzo Rovigatti,[b] Tommaso Bellini[c] and Francesco Sciortino[d]

Received 11th April 2012, Accepted 24th May 2012
DOI: 10.1039/c2sm25845e

Short blunt-ended DNA duplexes comprising 6 to 20 base pairs self-assemble into polydisperse semi-flexible chains due to hydrophobic stacking interactions between terminal base pairs. Above a critical concentration, which depends on temperature and duplex length, such chains order into liquid crystal phases. Here, we investigate the self-assembly of such double-helical duplexes with a combined numerical and theoretical approach. We simulate the bulk system employing the coarse-grained DNA model recently proposed by Ouldridge *et al.* [*J. Chem. Phys.*, 2011, **134**, 08501]. Then we evaluate the input quantities for the theoretical framework directly from the DNA model. The resulting parameter-free theoretical predictions provide an accurate description of the simulation results in the isotropic phase and theoretical values for the isotropic–nematic phase boundaries which are in line with experimental findings. In addition, the developed theoretical framework makes it possible to provide a route to estimate the stacking free energy.

1 Introduction

Self-assembly is the spontaneous formation through free energy minimization of reversible aggregates of basic building blocks. The size of the aggregating units, *e.g.* simple molecules, macromolecules or colloidal particles, can vary from a few angströms to microns, thus making self-assembly ubiquitous in nature and of interest in several fields, including material science, soft matter and biophysics.[1–5] Through self-assembly it is possible to design new materials whose physical properties are controlled by tuning the interactions of the individual building blocks.[6–9]

A relevant self-assembly process is the formation of filamentous aggregates (*i.e.* linear chains) induced by the anisotropy of attractive interactions. Examples are provided by micellar systems,[10–12] formation of fibers and fibrils,[13–16] solutions of long duplex B-form DNA composed of 10^2 to 10^6 base pairs,[17–20] filamentous viruses,[21–25] chromonic liquid crystals[26] as well as inorganic nanoparticles.[27]

If linear aggregates possess sufficient rigidity, the system may exhibit liquid crystal (LC) phases (*e.g.* nematic or columnar) above a critical concentration. In the present study we focus on the self-assembly of short (*i.e.* 6 to 20 base pairs) DNA duplexes (DNADs)[28–30] in which coaxial stacking interactions between the blunt ends of the DNADs favor their aggregation into weakly bonded chains. Such a reversible physical polymerization is enough to promote the mutual alignment of these chains and the formation of macroscopic orientationally-ordered nematic LC phases. At present, stacking is understood in terms of hydrophobic forces acting between the flat hydrocarbon surfaces provided by the paired nucleobases at the duplex terminals, although the debate on the physical origin of such interactions is still active and lively.[31,32] In this respect, the self-assembly of DNA duplexes provides a suitable way to access and quantify hydrophobic coaxial stacking interactions.

In order to extract quantitative information from DNA–DNA coaxial stacking experiments, reliable computational models and theoretical frameworks are needed. Recent theoretical approaches have focused on the isotropic–nematic (I–N) transition in self-assembling systems,[33,34] building on previous work on rigid and semi-flexible polymers.[35–43] In a recent publication[44] we investigated the reversible physical polymerization and collective ordering of DNA duplexes by modeling them as super-quadrics with quasi-cylindrical shape[45] with two reactive sites[46,47] on their bases. Then we presented a theoretical framework, built on the theories of Wertheim[48–50] and Onsager,[51] which is able to properly account for the association process.

Here, we employ this theoretical framework to study the physical properties of a realistic coarse-grained model of DNA recently proposed by Ouldridge *et al.*,[52] where nucleotides are modeled as rigid bodies interacting with site–site potentials. Following ref. 44, we compute the inputs required by the theory, *i.e.* the stacking free energy and the DNAD excluded volume, for the model of Ouldridge *et al.*[52] Subsequently we predict the

[a]*Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, Roma, I-00185, Italy. E-mail: cristiano.demichele@roma1.infn.it; Fax: +39 06463158; Tel: +39 0649913524*
[b]*Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, Roma, I-00185, Italy*
[c]*Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università di Milano, I-20122 Milano, Italy*
[d]*Dipartimento di Fisica, and CNR-ISC, "Sapienza" Università di Roma, P.le A. Moro 2, 00185 Roma, Italy*

polymerization extent in the isotropic phase as well as the isotropic–nematic phase boundaries.

To validate the theoretical predictions, we perform large-scale molecular dynamics (MD) simulations in the NVT ensemble of a bulk system comprising 9600 nucleotides, a study made possible by the computational power of modern Graphical Processing Units (GPUs). The parameter-free theoretical predictions provide an accurate description of the simulation results in the isotropic phase, supporting the theoretical approach and its application in the comparison with experimental results.

The article is organized as follows. Section 2 provides details of the coarse-grained model of DNADs and of the MD computer simulations; Section 3 presents a summary of the theory. Section 4 describes the protocols implemented to evaluate the input parameters required by the theory *via* MC integrations for two DNADs. We also discuss some geometrical properties of the bonded dimer configurations. We then compare the theoretical predictions with simulation and experimental results. Finally, in Section 5 we discuss estimates for the stacking free energy and present our conclusions.

2 Methods

2.1 Model

We implement a coarse-grained model for DNA recently developed by Ouldridge *et al.*[52,53] In such a model, designed *via* a top-down approach, each nucleotide is described as a rigid body (see Fig. 1). The interaction forms and parameters are chosen so as to reproduce structural and thermodynamic properties of both single- (ssDNA) and double- (dsDNA) stranded molecules of DNA in B-form. All interactions between nucleotides are pairwise and, in the last version of the model,[52] continuous and differentiable. Such feature makes the model convenient for MD simulations.

The interactions between nucleotides account for excluded volume, backbone connectivity, Watson–Crick hydrogen bonding, stacking, cross-stacking and coaxial-stacking. The interaction parameters have been adjusted in order to be consistent with experimental data.[52,54,55] In particular, the stacking interaction strength and stiffness have been chosen to be consistent with the experimental results reported for 14-base oligomers by Holbrook *et al.*[55] Hydrogen-bonding and cross-stacking potentials were adjusted to give duplex and hairpin formation thermodynamics consistent with the SantaLucia parameterization of the nearest-neighbor model.[54] Interaction stiffnesses were also further adjusted in order to provide correct mechanical properties of the model, such as the persistence length. The model does not have any sequence dependence apart from the Watson–Crick pairing, meaning that the strength of the interactions acting between nucleotides is to be considered as an average value. In addition, the model assumes conditions of high salt molarity (0.5 M). In this model, the coaxial-stacking interaction acts between any two non-bonded nucleotides and is responsible for the duplex–duplex bonding. It has been parametrized[56] to reproduce experimental data which quantify the stacking interactions by observing the difference in the relative mobility of a double strand where one of the two strands has been nicked with respect to intact DNA[57,58] and by analyzing the melting temperatures of short duplexes adjacent to hairpins.[59]

To cope with the complexity of the model and the large number of nucleotides involved in bulk simulations, we employ a modified version of the CPU-GPU code used in a previous work,[60] and extend it to support the force-fields.[61] Harvesting the power of modern Graphical Processing Units (GPUs) results in a 30-fold speed-up. The CPU version of the code, as well as the Python library written to simplify generation of initial configurations and post-processing analysis have been released as free software.[62]

2.2 Bulk simulations

To compare numerical results with theory, we perform Brownian dynamics simulations of 400 dsDNA molecules made up of 24

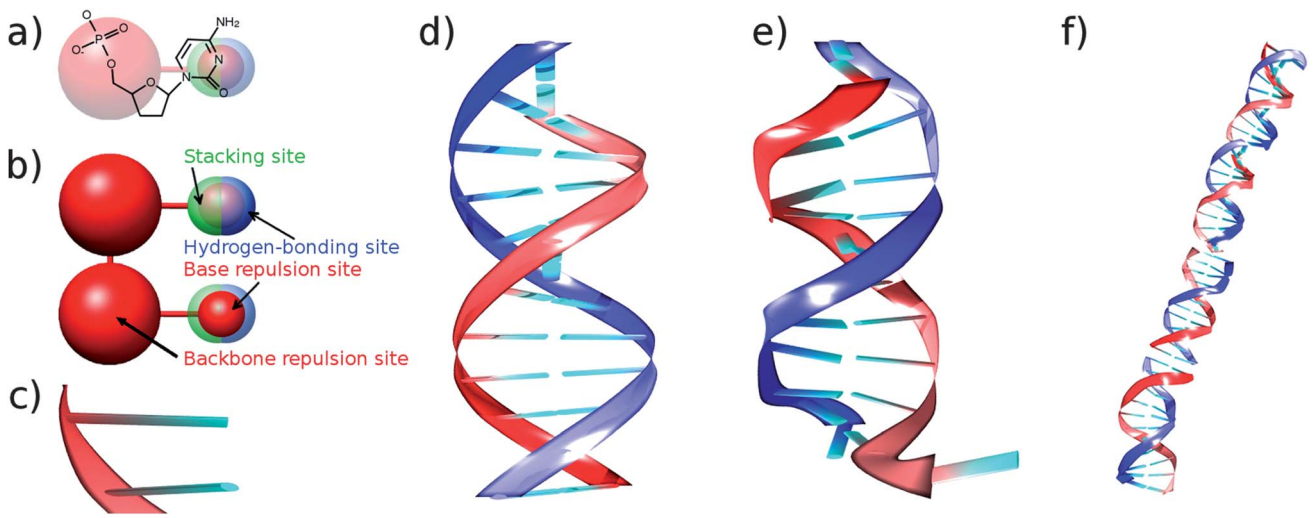

Fig. 1 (a) Schematic representation of the coarse graining of the model for a single nucleotide. (b) Model interaction sites. For the sake of clarity, stacking and hydrogen-bonding sites are highlighted on one nucleotide and the base repulsion site on the other. For visualization reasons, in the following strands will be shown as ribbons and bases as extended plates as depicted in (c). (d) A 12 base-pairs DNAD in the minimum energy configuration. (e) An equilibrium configuration of the same object at $T = 300$ K. The nucleotides at the bottom are not bonded, the so-called fraying effect. (f) A chain of length $N_b = 48$ extracted from a simulation at $c = 241$ mg ml^{-1} and $T = 270$ *K*.

nucleotides each, *i.e.* 400 cylinder-like objects with an aspect ratio of ≈2 (see Fig. 1(d)). The integration time step has been chosen to be 0.003 in simulation units which corresponds, if rescaled with the units of length, mass and energy used in the model, to approximately 1×10^{-14} seconds.

We study systems at three different temperatures, namely $T = 270$ K, 285 K and 300 K, and for different concentrations, ranging from 2 mg ml^{-1} to 241 mg ml^{-1}. The $T = 270$ K state point, despite being far from the experimentally accessed T, is here investigated to test the theory in a region of the phase diagram where the degree of association is significant. To quantify the aggregation process we define two DNADs as bonded if their pair interaction energy is negative. Depending on temperature and concentration, we use 10^6 to 10^7 MD steps for equilibration and 10^8 to 10^9 MD steps for data generation on NVIDIA Tesla C2050 GPUs, equivalent to 1–10 μs.

3 Theory

We build on the theoretical framework previously developed to account for the linear aggregation and collective ordering of quasi-cylindrical particles.[44] Here, we provide a discussion of how such a theory can be used to describe the reversible chaining and ordering of oligomeric DNADs at the level of detail adopted by the present model. According to ref. 44, the free energy of a system of equilibrium polymers can be written as

$$\frac{\beta F}{V} = \sum_{l=1}^{\infty} \nu(l)\{\ln[v_d \nu(l)] - 1\} + \frac{\eta(\phi)}{2} \sum_{\substack{l=1 \\ l'=1}}^{\infty} \nu(l)\nu(l')\, v_{excl}(l, l') - \beta \Delta F_b \sum_{l=1}^{\infty} (l-1)\nu(l) + \sum_{l=1}^{\infty} \nu(l)\sigma_o(l) \quad (1)$$

where V is the volume of the system, v_d is the volume of a monomer, $\phi \equiv v_d\rho$ ($\rho = N/V$ is the number density of monomers) is the packing fraction, $\nu(l)$ is the number density of chains of length l, normalized such that $\sum_{l=1}^{\infty} l\nu(l) = \rho$, ΔF_b, as discussed in Section 4.2, is a parameter which depends on the free energy associated to a single bond and $v_{excl}(l, l')$ is the excluded volume of two chains of length l and l'. $\eta(\phi)$ is the Parsons–Lee factor[63]

$$\eta(\phi) = \frac{1}{4}\frac{4 - 3\phi}{(1-\phi)^2} \quad (2)$$

and $\sigma_o(l)$[43] accounts for the orientational entropy that a chain of length l loses in the nematic phase (including possible contributions due to its flexibility). The explicit form for $\sigma_o(l)$ can be found in ref. 44.

The free energy functional (eqn (1)) explicitly accounts for the polydispersity inherent in the equilibrium polymerization process using a discrete chain length distribution and for the entropic and energetic contributions of each single bond through the parameter ΔF_b.

3.1 Isotropic phase

In the isotropic phase, $\sigma_o = 0$ and the excluded volume can be written as follows (see Appendix A):

$$v_{excl}(l, l', X_0) = 2B_I X_0^2 ll' + 2v_d k_I \frac{l + l'}{2} \quad (3)$$

where the parameters B_I and k_I can be estimated *via* MC integrals of a system composed of only two monomers (see Appendix A) and X_0 is the aspect ratio of the monomers. We assume that the chain length distribution $\nu(l)$ is exponential[44] with an average chain length M

$$\nu(l) = \rho M^{-(l+1)}(M-1)^{l-1} \quad (4)$$

where

$$M = \frac{\sum_{l=1}^{\infty} l\nu(l)}{\sum_{l=1}^{\infty} \nu(l)}. \quad (5)$$

With this choice for $\nu(l)$ the free energy in eqn (1) becomes:

$$\frac{\beta F_I}{V} = -\rho\beta\Delta F_b(1 - M^{-1}) + \eta(\phi)\left[B_I X_0^2 + \frac{v_d k_I}{M}\right]\rho^2 + \frac{\rho}{M}\left[\ln\left(\frac{v_d\rho}{M}\right) - 1\right] + \rho\frac{M-1}{M}\ln(M-1) - \rho\ln M. \quad (6)$$

Minimization of the free energy in eqn (6) with respect to M provides the following expression for the average chain length M:

$$M = \frac{1}{2}\left(1 + \sqrt{1 + 4\phi e^{k_I\phi\eta(\phi) + \beta\Delta F_b}}\right). \quad (7)$$

3.2 Nematic phase

In the nematic phase the monomer orientational distribution function $f(\theta)$ depends explicitly on the angle θ between the particle and the nematic axis, *i.e.* the direction of average orientation of the DNAD, since the system is supposed to have azimuthal symmetry around such an axis. We assume the form proposed by Onsager,[51] *i.e.*:

$$f_\alpha(\theta) = \frac{\alpha}{4\pi \sinh\alpha}\cosh(\alpha\cos\theta) \quad (8)$$

where α controls the width of the angular distribution. Its equilibrium value is obtained by minimizing the free energy with respect to α. As discussed in Appendix A, we assume the following form for the excluded volume in the nematic phase:

$$v_{excl}(l, l', X_0, \alpha) = 2B_N(\alpha)X_0^2 ll' + 2v_d k_N^{HC}(\alpha)\frac{l + l'}{2} \quad (9)$$

where the term $2v_d k_N^{HC}(\alpha)$ is the end–midsection contribution to the excluded volume of two hard cylinders (see Appendix B) and

$$B_N(\alpha) = \frac{\pi}{4}D^3\left(\eta_1 + \frac{\eta_2}{\alpha^{1/2}} + \frac{\eta_3}{\alpha}\right) \quad (10)$$

where D is the diameter of the monomer and η_k with $k = 1, 2, 3$ are three parameters that we chose in order to reproduce the excluded volume calculated from MC calculations as discussed in Appendix A.

Inserting eqn (9) and (4) into eqn (1), and assuming once more an exponential distribution for $\nu(l)$, one obtains after some algebra:

$$\frac{\beta F_N}{V} = \hat{\sigma}_o - \rho\beta\Delta F_b(1 - M^{-1}) + \eta(\phi)\left[B_N(\alpha)X_0^2 + \frac{v_d k_N^{HC}(\alpha)}{M}\right]\rho^2 + \frac{\rho}{M}\left(\ln\left[\frac{v_d\rho}{M}\right] - 1\right) - \rho\ln M + \rho\ln(M-1)\frac{M-1}{M} \quad (11)$$

where $\hat{\sigma}_o \equiv \sum_l \sigma_o(l)\nu(l)$. The explicit calculation of the parameters B_N and k_N^{HC} is explained in Appendices A and B.

Assuming that the orientational entropy $\hat{\sigma}_o$ can be approximated with the expression valid for long chains,[43] minimization with respect to M results in

$$M = \frac{1}{2}\left(1 + \sqrt{1 + \alpha\phi e^{k_N(\alpha)\phi\eta(\phi)+\beta\Delta F_b}}\right). \quad (12)$$

while using the approximated expression for short chains,[43] one obtains

$$M = \frac{1}{2}\left(1 + \sqrt{1 + 4\alpha\phi e^{k_N(\alpha)\phi\eta(\phi)+\beta\Delta F_b - 1}}\right). \quad (13)$$

The equilibrium value of α is thus determined by further minimizing the nematic free energy in eqn (11), which has become only a function of α. The parameter α is related to the degree of orientational ordering in the nematic phase as expressed by the nematic order parameter S as follows:

$$S(\alpha) = \int(3\cos^2\theta - 1)f_\alpha(\theta)\pi\sin\theta\, d\theta \approx 1 - 3/\alpha. \quad (14)$$

Further refinements of the theory can be obtained by including a more accurate description of the orientational distribution $f_\alpha(\theta)$ in the proximity of the I–N phase transition, along the lines of eqn (40)–(42) of ref. 44. For the sake of simplicity we have just presented the basic theoretical treatment. However, in the theoretical calculations in Section 4 we will make use of the refined and more accurate free energy functional proposed in ref. 44.

3.3 Phase coexistence

The phase boundaries, at which the aggregates of DNAD are sufficiently long to induce macroscopic orientational ordering, are characterized by coexisting isotropic and nematic phases in which the volume fraction of DNADs are, respectively, $\phi_N = v_d\rho_N$ and $\phi_I = v_d\rho_I$. The number densities ρ_I and ρ_N can be calculated by minimizing eqn (6) with respect to M_I and by minimizing eqn (11) with respect to M_N and α. In addition, the two phases must be at equal pressure, *i.e.* $P_I = P_N$, and chemical potential, *i.e.* $\mu_I = \mu_N$. These conditions yield the following set of equations:

$$\begin{aligned}
&\frac{\partial}{\partial M_I}F_I(\rho_I, M_I) = 0\\
&\frac{\partial}{\partial M_N}F_N(\rho_N, M_N, \alpha) = 0\\
&\frac{\partial}{\partial \alpha}F_N(\rho_N, M_N, \alpha) = 0\\
&P_I(\rho_I, M_I) = P_N(\rho_N, M_N, \alpha)\\
&\mu_I(\rho_I, M_I) = \mu_N(\rho_N, M_N, \alpha)
\end{aligned} \quad (15)$$

4 Results and discussion

4.1 Properties of the model

To characterize structural and geometrical properties of monomers and aggregates, we analyze conformations of duplexes extracted from large-scale GPU simulations (see Fig. 2 for some snapshots).

In the following, the volume v_d occupied by a single DNAD of length X_0D and double helix diameter D ($D \simeq 2$ nm) will be considered as the volume of a cylinder with the same length and diameter, *i.e.* $v_d = \pi X_0 D^3/4$. When comparing numerical and experimental results with theoretical predictions we use the number of base pairs N_b in place of X_0 ($X_0 \simeq 0.172N_b$) and the concentration c instead of the packing fraction ϕ, which can be related to the former *via*:

$$\phi = \frac{0.172D^3\pi}{8m_N}c \quad (16)$$

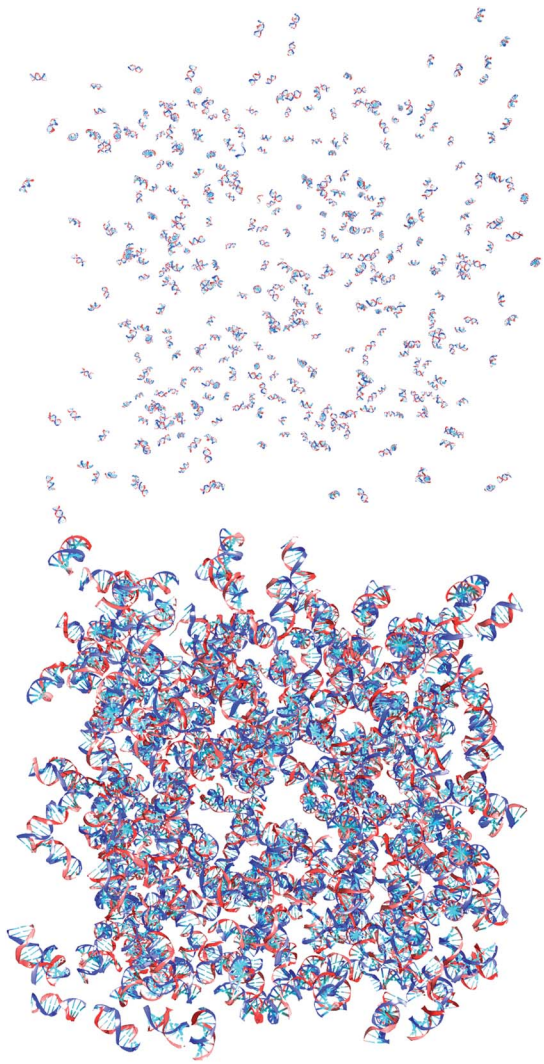

Fig. 2 Snapshots taken from simulations at $T = 300$ K. At low concentrations ($c = 2$ mg ml^{-1}, top) chain formation is negligible and the average chain length is approximately 1. As the concentration is increased ($c = 80$ mg ml^{-1}, bottom), DNADs start to self-assemble into chains and the average chain length increases.

where $m_N = 330$ Da is the average mass of a nucleotide. Hence, in the following c_I and c_N will be used in place of ϕ_I and ϕ_N.

First we calculate the dimensions (height L and width D) of the DNADs for different c and T. We observe no concentration dependence on both quantities, while the variation in T is negligible (of the order of 0.1% between DNADs of samples at 270 K and 300 K). The effect of this small change does not substantially affect the value of the aspect ratio, which we consider constant ($X_0 = 2.06$) throughout this work.

The geometrical properties of end-to-end bonded duplexes are not well-known since there are no experimental ways to probe such structures. In a very recent work, the interaction between duplex terminal base pairs has been analyzed by means of large-scale full-atom simulations by Maffeo *et al.*[64] They found that blunt-ended duplexes (*i.e.* duplexes without dangling ends) have preferential binding conformations with different values of the azimuthal angle γ, defined as the angle between the projections onto the plane orthogonal to the axis of the double helix of the vectors connecting the O5′ and O3′ terminal base pairs. They report two preferential values for γ, namely $\gamma = -20°$ and $\gamma = 180°$.

In the present model the continuity of the helix under end-to-end interactions is intrinsic in the model and the azimuthal angle probability distribution is peaked around a single value $\gamma_0 \approx 40°$ (see Fig. 3(a)). This is very close to the theoretical value $\gamma \approx 36°$ given by the pitch of the B-DNA double helix. The qualitative difference between conformations of bonded DNADs found in this work and in ref. 64 should be addressed in future studies, in order to describe the coaxial end-to-end interaction in a more proper way.

In addition, we calculate the average distance r between the centres of mass of the terminal base pairs. Fig. 3(b) shows $P(r)$, the probability distribution of r. $P(r)$ is peaked at 0.39 nm, whereas Maffeo *et al.*[64] found an average distance of $r \approx 0.5$ nm. This difference can be understood in terms of the effect of the salt concentration which, being five times higher than the one used in ref. 64, increases the electrostatic screening, thus effectively lowering the repulsion between DNA strands.

The effect of the temperature is small, as lowering T leads only to more peaked distributions for both $P(\gamma)$ and $P(r)$ (and a very small shift towards smaller angles for γ) but does not change the overall behavior.

4.2 Stacking free energy and excluded volume

In this section we discuss the procedure employed to evaluate the input quantities required by the theory, namely ΔF_b and $v_{excl}(l, l')$. To this aim we perform a Monte Carlo integration over the degrees of freedom of two duplexes. ΔF_b is defined as[44]

$$\beta \Delta F_b = \ln\left[2\frac{\Delta(T)}{v_d}\right] \tag{17}$$

where[65]

$$\Delta(T) = \frac{1}{4}\left\langle \int_{V_b} \left[e^{-\beta V(r_{12},\Omega_1,\Omega_2)} - 1\right] d\boldsymbol{r}_{12} \right\rangle. \tag{18}$$

Here $\boldsymbol{r}_{12}$ is the vector joining the center of mass of particles 1 and 2, Ω_i is the orientation of particle i and $\langle \ldots \rangle$ represents an average taken over all possible orientations. V_b is the bonding volume, defined here as the set of points where the interaction energy $V(\boldsymbol{r}_{12}, \Omega_1, \Omega_2)$ between duplex 1 and duplex 2 is less than $k_B T$. To numerically evaluate $\Delta(T)$ we perform a MC integration using the following scheme:

(1) Produce an ensemble of 500 equilibrium configurations of a single duplex at temperature T.

(2) Set the counter $N_{tries} = 0$ and the energy factor $F = 0$.

(3) Choose randomly two configurations i and j from the generated ensemble.

(4) Insert a randomly oriented duplex i in a random position in a cubic box of volume $V = 1000$ nm^3. Insert a second duplex j in a random position and with a random orientation. Compute the interaction energy $V(i, j)$ between the two duplexes i and j and, if $V(i, j) < k_B T$, update the energy factor, $F = F + (e^{-\beta V(i,j)} - 1)$. Increment N_{tries}.

(5) Repeat from step 3, until $\Delta(T) \cong \frac{1}{4}\frac{V}{N_{tries}}F$ converges within a few per cent precision.

The procedure employed to compute $v_{excl}(l, l')$ is fairly similar except that it is performed for duplexes with a various number of bases (*i.e.* with different X_0) and the quantity F counts how many trials originate from a pair configuration with $V(i, j) > k_B T$ (*i.e.* in step 4, $F = F + 1$). In the nematic case, the orientations of the duplexes are extracted randomly from the Onsager distribution given by eqn (8). With such procedure,

$$v_{excl}(l = 1, l' = 1, X_0) = \frac{V}{N_{tries}}F \tag{19}$$

We calculate v_{excl} for 8 values of α, ranging from 5 to 45 (see Appendix A). Since the X_0 and l dependences of eqn (3) and (9)

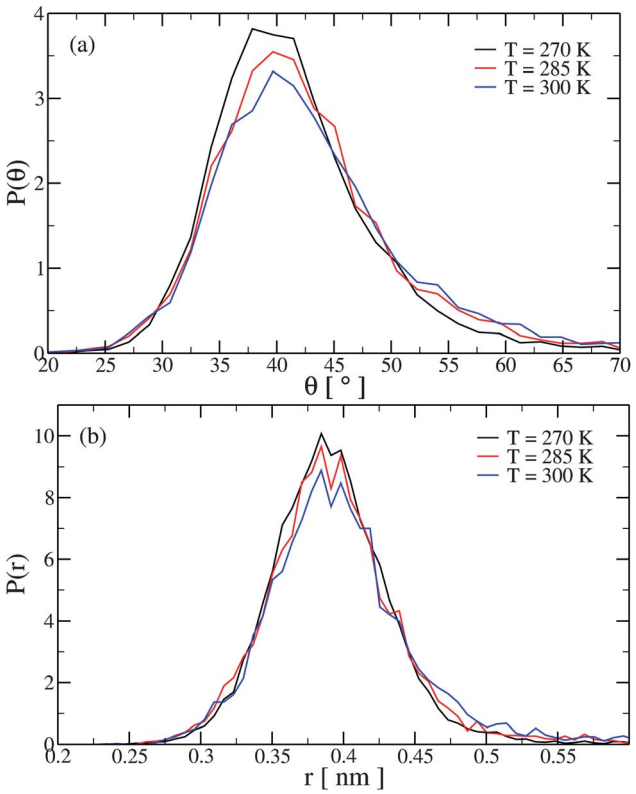

Fig. 3 Probability distributions for (a) the azimuthal angle γ and (b) the end-to-end distance r.

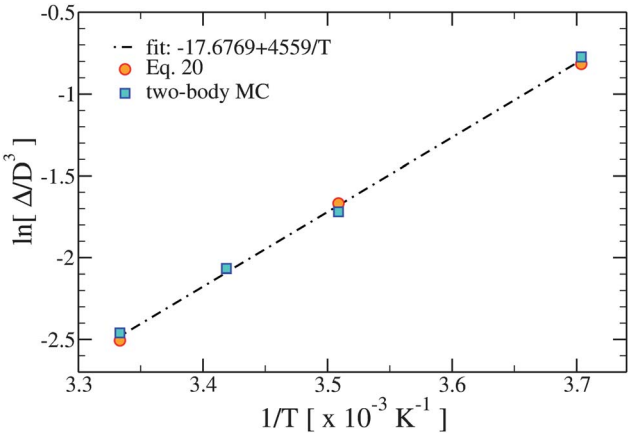

Fig. 4 $\Delta(T)$ calculated with the procedures described in Section 4.2 for all investigated T.

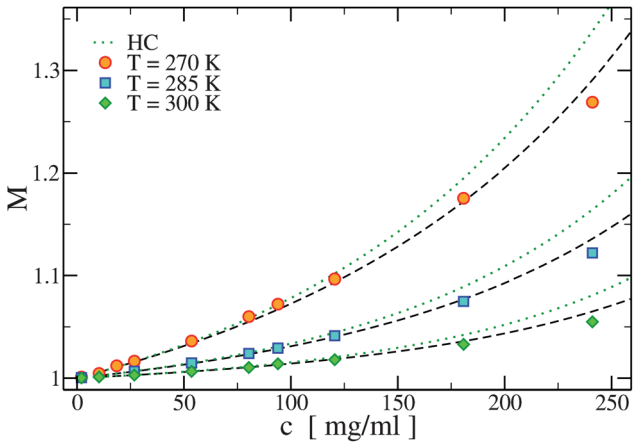

Fig. 5 Average chain length M in the isotropic phase at low concentration. Symbols are numerical results and dashed lines are theoretical predictions. Dotted lines are theoretical predictions using the excluded volume of HCs v_{excl}^{HC} (see Appendix B).

are the same and the X_0 dependence of the numerically calculated v_{excl} on the shape of DNADs is negligible, the evaluation of the excluded volume as a function of X_0 provides the same information as the evaluation of v_{excl} as a function of l.

We have checked that the dependence of $\Delta(T)$ and $v_{excl}(l, l')$ on the energy threshold employed in step 4 is negligible.

Fig. 4 shows $\Delta(T)$ for all investigated T in a $\ln \Delta$ *vs.* $1/T$ plot. A linear dependence properly describes the data at the three T. An alternative way to evaluate $\Delta(T)$ is provided by the limit $\rho \to 0$ of eqn (7). Indeed in the low density limit M and $\Delta(T)$ are related *via* the following relation:

$$\Delta(T) = \frac{M(1-M)}{2\rho}. \tag{20}$$

Therefore it is also possible to estimate $\Delta(T)$ by extrapolating the low density data for M at $T = 270$ K, 285 K and 300 K. The results, also shown in Fig. 4, are in line with the ones obtained through MC calculations. The Arrhenius behavior of $\Delta(T)$ suggests that bonding entropy and stacking energy are in first approximation T independent. The coaxial stacking free energy G_{ST} is related to $\Delta(T)$ as follows

$$G_{ST} = -k_B T \ln[2\rho\Delta(T)] \tag{21}$$

Substituting the fit expression provided in Fig. 4 for $\Delta(T)$ results in a stacking free energy $G_{ST}^0 = -0.086$ kcal mol^{-1} at a standard concentration 1 M of DNADs and $T = 293$ K comprising a bonding entropy of -30.6 cal mol^{-1} K^{-1} and a bonding energy of -9.06 kcal mol^{-1}.

4.3 Isotropic phase: comparing simulation results with theoretical predictions

Fig. 5 shows the concentration c dependence of M calculated from the MD simulation of the $N_b = 12$ system. The average chain length increases progressively on increasing c. The figure also shows the theoretical predictions calculated by minimizing the isotropic free energy in eqn (6) with respect to M using the previously discussed estimates for ΔF_b and v_{excl}. The theoretical results properly describe the MD simulation data up to concentrations around 200 mg ml^{-1}, which corresponds to a volume fraction $\phi \approx 0.20$. In ref. 44 similar observations have been made and the discrepancy at moderate and high ϕ has been attributed to the inaccuracy of the Parsons decoupling approximation. The M values calculated using the excluded volume of two hard cylinders (HC) are also reported to quantify the relevance of the actual shape of the DNA duplex. Indeed the HC predictions appreciably deviate from numerical data beyond 100 mg ml^{-1}.

4.4 Phase coexistence: theoretical predictions

A numerical evaluation of the phase coexistence between the isotropic and the nematic phases for the coarse-grained model adopted in this study is still impossible to obtain given the current computational power. We thus limit ourselves to the evaluation of the I–N phase coexistence *via* the theoretical approach discussed in Section 3. Fig. 6 shows the theoretical phase diagram in the c–N_b plane for $T = 270$ K and 300 K. As expected, both c_I and c_N decrease on increasing N_b, since the increase of the number of bases results in a larger aspect ratio. On decreasing T, theory predicts a 10% decrease of c_I and a similar decrease of c_N, resulting in an overall shift of the I–N coexistence region toward lower c values. This trend is related to the increase of the average chain length M with increasing $\beta\Delta F_b$ (see Fig. 4).

Fig. 6 also shows the phase boundaries calculated using the excluded volume of two hard cylinders. Assimilating DNADs to hard cylinders results in a 10–15% widening of the isotropic–nematic coexistence region.

4.5 Comparison between theory and experiments

The theoretical predictions concerning the isotropic–nematic coexisting concentrations can be compared to the experimental results reported in ref. 28 and 30 for blunt-ended DNADs.

Fig. 7 compares the experimentally determined nematic concentrations c_N at coexistence with the values calculated from the present model for $T = 293$ K. Despite all the simplifying assumptions and despite the experimental uncertainty, the results provide a reasonable description of the N_b dependence of c_N.

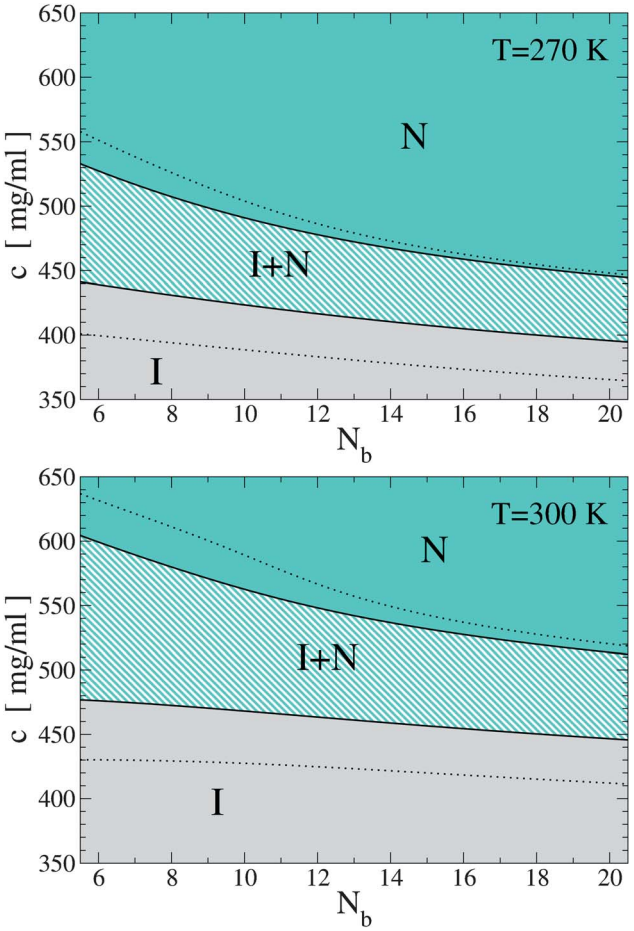

Fig. 6 I–N phase diagram in the c *vs.* N_b plane for $T = 270$ K (top) and 300 K (bottom). Dotted lines are theoretical phase boundaries calculated using the excluded volume of HCs v_{excl}^{HC} (see Appendix B).

The experimental data refer to different base sequences and different salt concentrations. According to the authors, c_N is affected by an error of about ± 50 mg ml^{-1}. In particular, for the case $N_b = 12$ the critical concentrations c_N for distinct sequences show that blunt-end duplexes of equal length but different sequences can display significantly different transition concentrations. Hence, for each duplex length, we consider the lowest transition concentration among the ones experimentally determined, since this corresponds to the sequence closest to the symmetric monomer in the model. Indeed the dependence of c_N on the DNADs sequence is expected to be larger for the shortest sequences, *i.e.* $N_b \leqq 12$, for which DNAD bending could be significant.[30] Unfortunately, quantitative experimental data on this bending effect are still lacking. In general it is possible that c_N, for $N_b < 12$ (for $N_b = 12$, where a large number of sequences have been studied, see Fig. 7), would be corrected to lower values if a larger number of sequences was explored. For more details on this phenomenon, we refer the reader to the discussions in ref. 30,44 and 66.

The overestimation of the phase boundaries for $N_b \geq 12$ with respect to experimental results suggests that the DNA model of Ouldridge *et al.*[52] overestimates the coaxial stacking free energy. Such a discrepancy can perhaps be attributed to the restricted number of microstates allowing for bonding states in the DNA model,[52,56] as discussed in Section 4. Indeed, allowing DNADs to form end-to-end bonds with more than one preferred azimuthal angle would increase the entropy of bonding, thus effectively lowering G_{ST}. Allowing for both left- and right-handed binding conformations, a possibility supported by the results of Maffeo *et al.*,[64] would double $\Delta(T)$ in eqn (21) and hence add an entropic contribution equal to $-k_B T \ln(2)$ to G_{ST}, which would result in a value $G_{ST}^0 - 0.403$ kcal mol^{-1} $= -0.49$ kcal mol^{-1} for $T = 293$ *K*. Fig. 7 also shows the theoretical predictions for such an upgraded G_{ST} value.

In Fig. 7 theoretical calculations of the I–N transition lines are shown for $G_{ST} = -0.4$ kcal mol^{-1} and $G_{ST} = -2.4$ kcal mol^{-1} at $T = 293$ K as the upper and lower boundaries of the grey band, respectively. To calculate these critical lines we retain the excluded volume calculated in Section 4.2 and, given the value of G_{ST}, we evaluate ΔF_b according to eqn (17) and (21) for $T = 293$ K and ρ corresponding to the standard 1 M concentration.

The selected points with $N_b \geq 12$ fall within the grey band shown in Fig. 7, enabling us to provide an indirect estimate of G_{ST} between -0.4 kcal mol^{-1} and -2.4 kcal mol^{-1}. For the points with $N_b < 12$, where duplex bending might play a role, it would be valuable to have more experimental points corresponding to more straight sequences in order to validate the theoretical predictions.

It is worth observing that for all DNAD lengths N_b, the electrostatic interactions are properly screened. For $N_b = 20$ a concentration 1.2 M of NaCl has been added to the solution resulting in a Debye screening length $k_D^{-1} \approx 0.23$ nm. For all other lengths (*i.e.* $N_b \leq 16$) we note that at the lowest DNA concentration of 440 mg ml^{-1} corresponding to $N_b = 14$ $k_D^{-1} \approx 0.40$ nm. Therefore the experimental k_D^{-1} is always smaller than the excluded volume diameter for the backbone–backbone interaction of our coarse-grained model[52] (≈ 0.6 nm), thus enabling us to neglect electrostatic interactions.

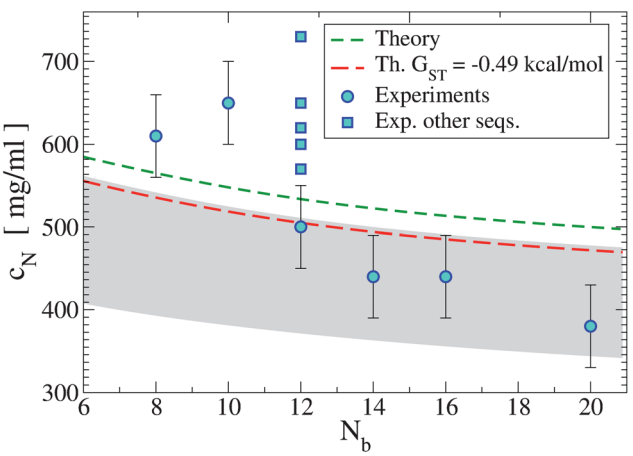

Fig. 7 Critical nematic concentrations c_N as a function of the number of base pairs per duplex N_b for the present model, calculated theoretically at $T = 293$ K using the computed stacking free energy G_{ST}^0 (short dashed lines), $G_{ST} = -0.49$ kcal mol^{-1} (long dashed lines), and for experiments[28] (circles and squares). Squares are c_N for different sequences at the same $N_b = 12$. The grey band has been built considering for G_{ST} an upper bound of -0.4 kcal mol^{-1} and a lower bound of -2.4 kcal mol^{-1}.

On the other hand, if electrostatic interactions are not properly screened the effective aspect ratio for such DNAD sequences would be smaller than the ones used in our theoretical treatment and this would result in an underestimate of c_N. To account for this behavior one should at least have a reasonable estimate of the effective size of DNADs when electrostatic interactions are not fully screened. Moreover, the role of electrostatics can be subtle and not completely accounted for by simply introducing an effective size of DNADs. A possible route to include electrostatics in our treatment can be found in ref. 20 and it will be addressed in future studies.

4.6 Comparison with Onsager theory

The experimental average aggregation numbers are estimated in ref. 13 and 28 by mapping the self-assembled system onto an "equivalent" mono-disperse system of hard rods with an aspect ratio equal to MX_0. In ref. 44 it has been shown that the theoretically estimated isotropic–nematic coexistence lines for the case of polymerizing superquadric particles in the MX_0–ϕ plane, parametrized by the stacking energy, are significantly different from the corresponding Onsager original predictions (as re-evaluated in ref. 35). In light of the relevance for interpreting the experimental data, we show in Fig. 8 the same curves for the DNA model investigated here. In this model, a clear re-entrant behavior of the transition lines in the c–MX_0 plane is observed. The re-entrant behavior occurs for values of the stacking free energy accessed at temperatures between 270 K and 330 K and it arises as a result of the competition between steric (entropic) and temperature (energetic) contributions to the free energy in driving the self-assembly process which leads the system to the isotropic–nematic transition. We believe that the re-entrance of the transition lines in the c–MX_0 plane is a peculiar mark of the system's polydispersity, which results from the reversible self-assembly of chains.

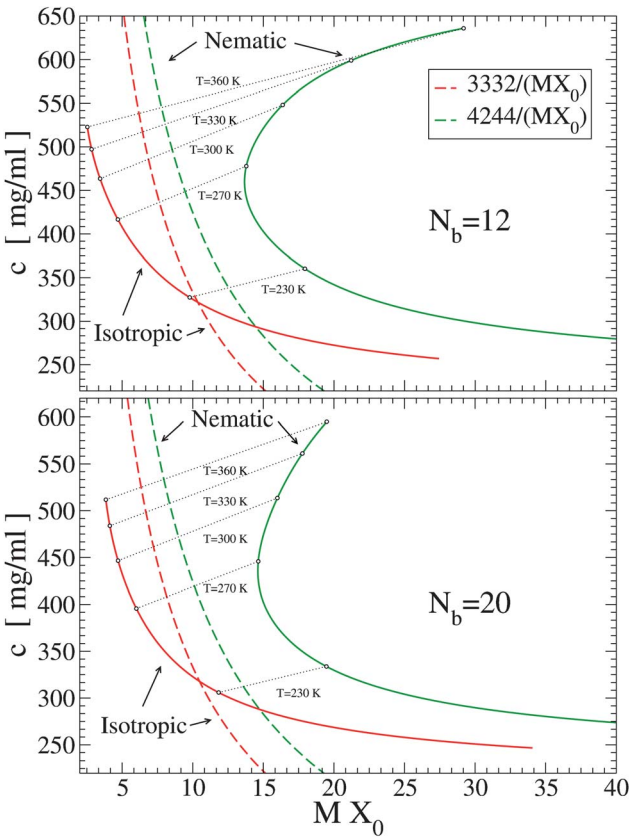

Fig. 8 Isotropic–nematic coexistence lines in the average aspect ratio MX_0 and concentration c plane for two values of N_b, namely $N_b = 12$ (top) and $N_b = 20$ (bottom). Solid lines indicate theoretical predictions, dashed lines indicate Onsager original predictions, as re-evaluated in ref. 34 for ϕ_I and ϕ_N and here reported in terms of the concentrations c_I and c_N expressed in mg ml^{-1}. Symbols along the isotropic and nematic phase boundaries at coexistence are joined by dotted lines, to indicate the change in concentration and average chain length at the transition.

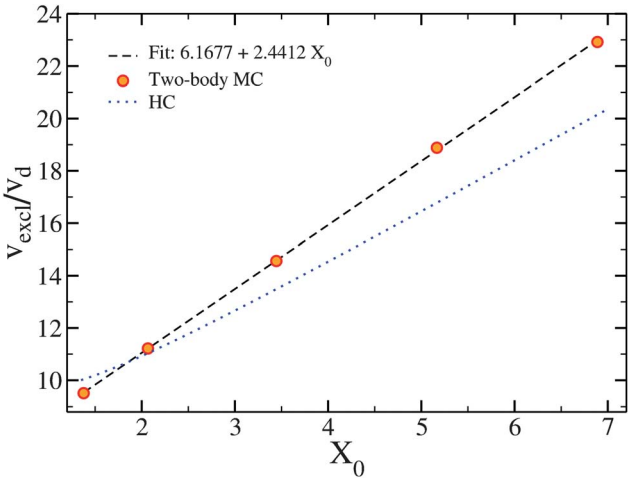

Fig. 9 Excluded volume in the isotropic phase together with analytic approximations. From the linear fit one has $B_I = 0.959D^3$ and $k_I = 3.084$, while we assume $A_I = 0$.

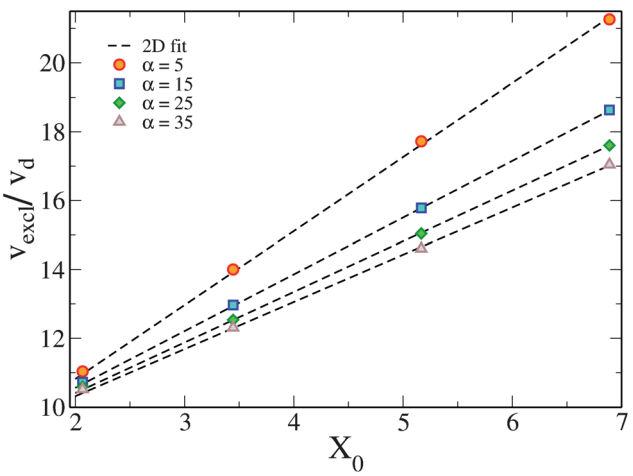

Fig. 10 Excluded volume as a function of aspect ratio X_0 in the nematic phase together with analytic approximations for several α. The dashed lines are obtained plotting the function reported in eqn (9) and setting $\eta_1 = 0.386419$, $\eta_2 = 1.91328$ and $\eta_3 = -0.836354$.

5 Conclusions

In this article we have studied an equilibrium bulk solution of blunt ended DNA duplexes undergoing reversible self-assembly into chains, promoted by stacking interactions. The simulation study, carried out at different concentrations and temperatures, provides a clear characterization of the c and T dependence of the average polymerization length M and an indirect estimate of the stacking free energy. We have provided a theoretical description of the self-assembly process based on a theoretical framework recently developed in ref. 44. The inputs required by the theory (the DNAD excluded volumes and the stacking free energy) have been numerically calculated for the present DNA model, allowing for a parameter free comparison between molecular dynamic results and theoretical predictions. Such a comparison has been limited to the isotropic phase, due to the difficulties in simulating the dense nematic phase under equilibrium conditions. The description of the isotropic phase is satisfactory: quantitative agreement between theory and simulations is achieved for concentrations up to $c \approx 200$ mg ml^{-1}. The stacking free energy value that properly accounts for the polymerization process observed in the molecular dynamics simulations is $G^0_{ST} = -0.086$ kcal mol^{-1} at a standard concentration 1 M of DNADs and $T = 293$ K comprising a bonding entropy of -30.6 cal mol^{-1} K^{-1} and a bonding energy of -9.06 kcal mol^{-1}.

Theoretical predictions for the I–N transition have been compared with experimental results for several DNA lengths, ranging from 8 to 20 bases. For $N_b \geq 12$, the model predicts values for c_N which are higher than experimental ones. This suggests that the DNA model employed overestimates G_{ST}. In view of the recent results of Maffeo *et al.*,[64] we speculate that the bonding entropy is underestimated, in agreement with the observation that the probability distribution of the azimuthal angle between two bonded DNADs, which is designed to be single-peaked, is too restraining. In this respect, the present study calls for an improvement of the coarse-grained potential[52] in regard to the coaxial stacking interaction.

The value of G_{ST} can also be used as a fitting parameter in the theory for matching c_N with the experimental results, retaining the excluded volume estimates calculated for the coarse-grained DNA model. Such a procedure shows that values of the stacking free energy between -0.4 kcal mol^{-1} and -2.4 kcal mol^{-1} are compatible with the experimental location of the I–N transition line. In the work of Maffeo *et al.*, the authors report a quite smaller value of G_{ST}, namely $G^M_{ST} = -6.3$ kcal mol^{-1}, a value which was confirmed by the same authors by performing an investigation of the aggregation kinetic in a very lengthy all-atom simulation of DNAD with $N_b = 10$. If such G_{ST} value is selected as input in our theoretical approach (maintaining the same excluded volume term), then one finds $c^M_N \approx 250$ mg ml^{-1}, a value significantly smaller than the experimental result ($c_N = 650 \pm 50$ mg ml^{-1}). This casts some doubts on the effectiveness of the employed all-atom force-field to properly model coaxial stacking.

Finally, our work draws attention to the errors affecting the estimate of the average chain length M *via* a straightforward comparison of the nematic coexisting concentrations with analytic predictions based on the original Onsager theory for mono-disperse thin rods.[13,28] We have found that such an approximation significantly underestimates M at the I–N transition concentration c_N. In addition, the theoretical approach predicts a re-entrant behavior of the transition lines in the c–MX_0 plane, a distinct feature of the polydisperse nature of the equilibrium chains.

Appendix A: excluded volume contributions

Here we further discuss the calculation of the excluded volume term $v_{excl}(l, l')$ for the present model. Following ref. 44, the excluded volume is assumed to be the following second order polynomial in l and l':

$$v_{excl}[l, l'; f(\theta)] = 2\int f(\theta)f(\theta')D^3\left[\Psi_1(\gamma, X_0) + \frac{l+l'}{2}\Psi_2(\gamma, X_0)X_0 + \Psi_3(\gamma, X_0)X_0^2 ll'\right]d\Omega d\Omega' \quad (22)$$

where the functions Ψ_α, $\alpha = 1, 2, 3$, describe the angular dependence of the excluded volume. The orientational probability $f(\theta)$ is normalized such that

$$\int f(\theta)d\Omega = 1 \quad (23)$$

The three contributions to the excluded volume in eqn (22) come from end–end, end–midsection and midsection–midsection steric interactions[44] between two chains.

In the isotropic phase the orientational distribution does not have any angular dependence, *i.e.* $f(\theta) = 1/4\pi$, and eqn (22) reduces to the form

$$v_{excl}(l, l', X_0) = B_I X_0^2 ll' + k_I v_d \frac{l+l'}{2} + A_I. \quad (24)$$

The parameters B_I, k_I and A_I appearing in eqn (24) can be calculated *via* MC integration procedures as discussed in Section 4.2 and in ref. 44. We expect that these parameters do not depend on X_0 because each DNAD comprises N_b stacked base pairs which are all identical with respect to excluded volume interactions (*i.e.* they all have the same shape). In particular, the calculated excluded volume of two DNADs is reported in Fig. 9 for 5 different aspect ratios, together with the resulting values for the above parameters.

Using the Onsager angular distribution $f_\alpha(\theta)$ in eqn (24), the excluded volume in the nematic phase depends also on the parameter α, *i.e.* the general form in eqn (22) reduces to

$$v_{excl}(l, l', X_0, \alpha) = B_N(\alpha)X_0^2 ll' + k_N(\alpha)v_d\frac{l+l'}{2} + A_N(\alpha). \quad (25)$$

Assuming that $A_N(\alpha) = 0$, $k_N(\alpha) = k^{HC}_N(\alpha)$ and $B_N(\alpha)$ is given by eqn (10), the three parameters η_k with $k = 1, 2, 3$ have to be estimated. For $l = l' = 1$ and several values of α ($\alpha = 5\ldots45$ in steps of 5) and X_0 we calculated numerically the nematic excluded volume for two DNADs. The results are shown in Fig. 10, where we plot v_{excl}/v_d *vs.* X_0 for various α. The dashed lines shown in Fig. 10 are obtained through a two-dimensional fit to numerical data for $v_{excl}(1, 1, X_0, \alpha)$ using eqn (9) as the fitting function.

Appendix B: excluded volume of hard cylinders

For two rigid chains of length l and l' which are composed of hard cylinders (HCs) of diameter D and length X_0D, $v_{\mathrm{excl}}(l, l')$ can be described by ref. 51.

$$v_{\mathrm{excl}}^{\mathrm{HC}}[l, l'; f(\theta)] = \int f(\theta) f(\theta') D^3 \Big[\frac{\pi}{2} \sin\gamma + \frac{\pi}{2} X_0 \Big(1 + |\cos\gamma| + \frac{4}{\pi} E(\sin\gamma) \Big) \frac{l + l'}{2} + 2X_0^2\, ll' \sin\gamma \Big] \mathrm{d}\Omega \mathrm{d}\Omega' \tag{26}$$

where $\cos\gamma = \boldsymbol{u}\cdot\boldsymbol{u}'$, $\boldsymbol{u}$ and $\boldsymbol{u}'$ are the orientations of two HCs and $E(\sin\gamma)$ is the complete elliptical integral.

$$E(\sin\gamma) = \frac{1}{4} \int_0^{2\pi} \left(1 - \sin^2\gamma \sin^2\psi\right)^{1/2} \mathrm{d}\psi. \tag{27}$$

The integrals in eqn (26) can be calculated exactly in the isotropic phase, while in the nematic phase the calculation can be done analytically only for suitable choices of the angular distribution $f(\theta)$. Here we assume that the angular distribution is given by the Onsager function in eqn (8).

Using the Onsager orientational function the following approximate expressions for the coefficients $k_{\mathrm{N}}(\alpha)$, $B_{\mathrm{N}}(\alpha)$ and $A_{\mathrm{N}}(\alpha)$ can be derived[43]

$$\begin{aligned} \tilde{B}_{\mathrm{N}}(\alpha) &= D^3(\pi/4)\rho_{\mathrm{a}}(\alpha) \\ \tilde{k}_{\mathrm{N}}(\alpha) &= \pi D^3 \frac{X_0}{v_{\mathrm{d}}} \left(1 - \frac{1}{\alpha}\right) \\ \tilde{A}_{\mathrm{N}}(\alpha) &= D^3(\pi/4)^2 \rho_{\mathrm{a}}(\alpha) \end{aligned} \tag{28}$$

where

$$\rho_{\mathrm{a}} = 4(\pi\alpha)^{-1/2} \left(1 - \frac{15}{16\alpha} + \frac{105}{512\alpha^2} + \frac{315}{8192\alpha^3}\right). \tag{29}$$

We evaluate numerically the excluded volume in eqn (26) for many values of α and, building on the expressions in eqn (28), we perform a fit to these data using the following functions:

$$B_{\mathrm{N}}^{\mathrm{HC}}(\alpha) \simeq D^3(\pi/4) \left(\rho_{\mathrm{a}}(\alpha) + \frac{c_4}{\alpha^{9/2}} + \frac{c_5}{\alpha^{11/2}}\right) \tag{30}$$

$$k_{\mathrm{N}}^{\mathrm{HC}}(\alpha) = 4\left(1 - \frac{1}{\alpha}\right) + \sum_{i=2}^{\infty} \frac{b_i}{\alpha^i} \simeq \frac{4}{\pi} \sum_{i=0}^{4} \frac{d_i}{\alpha^i} \tag{31}$$

$$A_{\mathrm{N}}^{\mathrm{HC}}(\alpha) \simeq D^3(\pi/4)^2 \left(\rho_{\mathrm{a}}(\alpha) + \frac{c_4}{\alpha^{9/2}} + \frac{c_5}{\alpha^{11/2}}\right) \tag{32}$$

The coefficient values resulting from the fitting procedure are $c_4 = 1.2563$, $c_5 = -0.95535$, $d_0 = 3.0846$, $d_1 = -4.0872$, $d_2 = 9.0137$, $d_3 = -9.009$ and $d_4 = 3.3461$.

Acknowledgements

We thank Giuseppe D'Adamo, Thomas Ouldridge, Flavio Romano and Teun Vissers for fruitful discussions. We acknowledge support from ERC-226207-PATCHYCOLLOIDS and ITN-234810-COMPLOIDS as well as from NVIDIA.

References

1 I. Hamley, *Introduction to Soft Matter*, Wiley & Sons, 2007.
2 S. C. Glotzer, *Science*, 2004, **306**, 419–420.
3 G. M. Whitesides and M. Boncheva, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 4769–4774.
4 Y.-S. Cho, G.-R. Yi, J.-M. Lim, S.-H. Kim, V. N. Manoharan, D. J. Pine and S.-M. Yang, *J. Am. Chem. Soc.*, 2005, **127**, 15968–15975.
5 F. W. Starr and F. Sciortino, *J. Phys.: Condens. Matter*, 2006, **18**, L347–L353.
6 Z. Nie, D. Fava, E. Kumacheva, S. Zou, G. C. Walker and M. Rubinstein, *Nat. Mater.*, 2007, **6**, 609–614.
7 J. P. K. Doye, A. A. Louis, I.-C. Lin, L. R. Allen, E. G. Noya, A. W. Wilber, H. C. Kok and R. Lyus, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2197–2205.
8 C. Mirkin, R. Letsinger, R. Mucic and J. Storhoff, *Nature*, 1996, **382**, 607.
9 V. Workum and J. Douglas, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **73**, 031502.
10 A. Khan, *Curr. Opin. Colloid Interface Sci.*, 1996, **1**, 614.
11 P. van der Schoot and M. Cates, *Langmuir*, 1994, **10**, 670–679.
12 D. M. Kuntz and L. M. Walker, *Soft Matter*, 2008, **4**, 286–293.
13 J.-M. Jung and R. Mezzenga, *Langmuir*, 2010, **26**, 504–514.
14 C. F. Lee, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2009, **80**, 031902.
15 A. Ciferri, *Liq. Cryst.*, 2007, **34**, 693–696.
16 A. Aggeli, M. Bell, L. M. Carrick, C. W. G. Fishwick, R. Harding, P. J. Mawer, S. E. Radford, A. E. Strong and N. Boden, *J. Am. Chem. Soc.*, 2003, **125**, 9619–9628.
17 C. Robinson, *Tetrahedron*, 1961, **13**, 219–234.
18 F. Livolant, A. M. Levelut, J. Doucet and J. P. Benoit, *Nature*, 1989, **339**, 724–726.
19 K. Merchant and R. L. Rill, *Biophys. J.*, 1997, **73**, 3154–3163.
20 F. Tombolato and A. Ferrarini, *J. Chem. Phys.*, 2005, **122**, 054908.
21 F. Tombolato, A. Ferrarini and E. Grelet, *Phys. Rev. Lett.*, 2006, **96**, 258302.
22 E. Barry, D. Beller and Z. Dogic, *Soft Matter*, 2009, **5**, 2563–2570.
23 E. Grelet and S. Fraden, *Phys. Rev. Lett.*, 2003, **90**, 198302.
24 S. Tomar, M. M. Green and L. A. Day, *J. Am. Chem. Soc.*, 2007, **129**, 3367–3375.
25 A. Minsky, E. Shimoni and D. Frenkiel-Krispin, *Nat. Rev. Mol. Cell Biol.*, 2002, **3**, 50–60.
26 J. Lydon, *J. Mater. Chem.*, 2010, **20**, 10071–10099.
27 K. Liu, Z. Nie, N. Zhao, W. Li, M. Rubinstein and E. Kumacheva, *Science*, 2010, **329**, 197–200.
28 M. Nakata, G. Zanchetta, B. D. Chapman, C. D. Jones, J. O. Cross, R. Pindak, T. Bellini and N. A. Clark, *Science*, 2007, **318**, 1276.
29 G. Zanchetta, M. Nakata, M. Buscaglia, N. A. Clark and T. Bellini, *J. Phys.: Condens. Matter*, 2008, **20**, 494214.
30 G. Zanchetta, F. Giavazzi, M. Nakata, M. Buscaglia, R. Cerbino, N. A. Clark and T. Bellini, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 17497–17502.
31 K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, D. C. Tahmassebi and E. T. Kool, *J. Am. Chem. Soc.*, 2000, **122**, 2213–2222.
32 T. Bellini, R. Cerbino and G. Zanchetta, *Liquid Crystals – Materials Design and Self-Assembly*, Springer, Berlin/Heidelberg, 2012, vol. 318, pp. 225–279.
33 T. Kuriabova, M. Betterton and M. Glaser, *J. Mater. Chem.*, 2010, **20**, 10366–10383.
34 X. Lü and J. Kindt, *J. Chem. Phys.*, 2004, **120**, 10328–10338.
35 G. J. Vroege and H. N. W. Lekkerkerker, *Rep. Prog. Phys.*, 1992, **55**, 1241–1309.
36 M. Dijkstra and D. Frenkel, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1995, **51**, 5891–5898.
37 A. Khokhlov and A. Semenov, *Phys. A*, 1981, **108**, 546–556.
38 A. Khokhlov and A. Semenov, *Phys. A*, 1982, **112**, 605–614.
39 P. P. F. Wessels and B. M. Mulder, *J. Phys.: Condens. Matter*, 2006, **18**, 9335.
40 M. Dennison, M. Dijkstra and R. van Roij, *Phys. Rev. Lett.*, 2011, **106**, 208302.
41 Z. Wang, D. Kuckling and D. Johannsmann, *Soft Matter*, 2003, **1**, 353–364.
42 Z. Y. Chen, *Macromolecules*, 1993, **26**, 3419–3423.

43 T. Odijk, *Macromolecules*, 1986, **19**, 2313–2329.
44 C. De Michele, T. Bellini and F. Sciortino, *Macromolecules*, 2012, **45**, 1090–1106.
45 C. De Michele, *J. Comput. Phys.*, 2010, **229**, 3276–3294.
46 F. Sciortino, C. De Michele, S. Corezzi, J. Russo, E. Zaccarelli and P. Tartaglia, *Soft Matter*, 2009, **5**, 2571–2575.
47 S. Corezzi, C. De Michele, E. Zaccarelli, P. Tartaglia and F. Sciortino, *J. Phys. Chem. B*, 2009, **113**, 1233–1236.
48 M. Wertheim, *J. Stat. Phys.*, 1984, **35**, 19–34.
49 M. Wertheim, *J. Stat. Phys.*, 1984, **35**, 35–47.
50 M. Wertheim, *J. Stat. Phys.*, 1986, **42**, 459–476.
51 L. Onsager, *Ann. N. Y. Acad. Sci.*, 1949, **51**, 627.
52 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, *J. Chem. Phys.*, 2011, **134**, 085101.
53 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, *Phys. Rev. Lett.*, 2010, **104**, 178101.
54 J. SantaLucia, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 1460.
55 J. A. Holbrook, M. W. Capp, R. M. Saecker and M. T. Record, *Biochemistry*, 1999, **38**, 8409–8422.
56 T. E. Ouldridge, Ph.D. thesis, University of Oxford, 2011.
57 E. Protozanova, P. Yakovchuk and M. Frank-Kamenetskii, *J. Mol. Biol.*, 2004, **342**, 775–785.
58 P. Yakovchuk, E. Protozanova and M. D. Frank-Kamenetskii, *Nucleic Acids Res.*, 2006, **34**, 564–574.
59 S. Bommarito, N. Peyret and J. J. SantaLucia, *Nucleic Acids Res.*, 2000, **28**, 1929–1934.
60 L. Rovigatti and F. Sciortino, *Mol. Phys.*, 2011, **109**, 2889–2896.
61 M. Allen and G. Germano, *Mol. Phys.*, 2006, **104**, 3225–3235.
62 P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, A. A. Louis and J. P. K. Doye, in preparation, http://dna.physics.ox.ac.uk.
63 J. Parsons, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1979, **19**, 1225–1230.
64 C. Maffeo, B. Luan and A. Aksimentiev, *Nucleic Acids Res.*, 2012, **40**, 3812–3821.
65 F. Sciortino, E. Bianchi, J. F. Douglas and P. Tartaglia, *J. Chem. Phys.*, 2007, **126**, 194903.
66 E. Frezza, F. Tombolato and A. Ferrarini, *Soft Matter*, 2011, **7**, 9291–9296.

