Soft Matter

PAPER

Self-assembly of mesogenic bent-core DNA nanoduplexes

Khanh Thuy Nguyen,[a] Anna Battisti,[a] Daniele Ancora,[a] Francesco Sciortino[b] and Cristiano De Michele*[c]

Short cylinder-like DNA duplexes, comprising 6 to 20 base pairs, self-assemble into semi-flexible chains, due to coaxial stacking interactions between their blunt ends. The mutual alignment of these chains gives rise to macroscopically orientationally ordered liquid crystal phases. Interestingly, experiments show that the isotropic–nematic phase boundary is sequence-dependent. We perform all-atom simulations of several sequences to gain insights into the structural properties of the duplex and correlate the resulting geometric properties with the observed location of the isotropic–nematic phase boundary. We identify in the duplex bending the key parameter for explaining the sequence dependence, suggesting that DNA duplexes can be assimilated to bent-core mesogens. We also develop a coarse-grained model for the different DNA duplexes to evaluate in detail how bending affects the persistence length and excluded volume of the aggregates. This information is fed into a recently developed formalism to predict the isotropic–nematic phase boundary for bent-core mesogens. The theoretical results agree with the experimental observations.

1. Introduction

Since the pioneering work of Nadrian Seeman,[1,2] which paved the way for DNA-based nanotechnology, DNA has been widely used in Soft Matter physics to design and build new materials.[3–14] DNA-based particles can be used as basic building blocks which self-assemble into reversible aggregates, giving rise to ordered,[15–18] partially ordered[19] or dynamically arrested disordered[20–22] phases under suitable conditions of temperature and concentrations.

A particular, but very interesting, case of DNA self-assembly emerged recently from a series of experiments[23–26] which have provided evidence that a solution of short DNA duplexes (DNAD), 6 to 20 base pairs in length, can form liquid crystals (LCs) above a critical concentration, giving rise to nematic and columnar LC phases.[23] Here, hydrophobic stacking interactions lead to the formation of semi-flexible chains. This behavior does not only pertain to B-form DNA oligomers since it has also been observed in solutions of blunt-ended A-form RNA oligomeric duplexes.[27]

Depending on the thermodynamic state point (temperature T and concentration c) these chains may attain the required anisotropy to undergo an isotropic to nematic transition (see Fig. 1). Further experiments[25] have also provided evidence that for ≤ 12 base pairs, the critical concentration is sequence dependent, suggesting a structural origin for this phenomenon.[28–30] It is tempting to speculate that short DNADs can be considered as bent-core molecules which could provide the rich and very interesting phenomenology of banana-shaped mesogens.[31] Recently, bent-core mesogens have received much attention from the liquid crystal community from theoretical, numerical[32–35] and experimental points of view.[36–38] The relevance is rooted in the work of Niori *et al.*[39] where achiral bent-

a Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, 00185 Roma, Italy

b ISC-CNR and Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, 00185 Roma, Italy

c ISC-CNR and Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, 00185 Roma, Italy. E-mail: cristiano.demichele@roma1.infn.it; Fax: +39 064463158; Tel: +39 0649913524
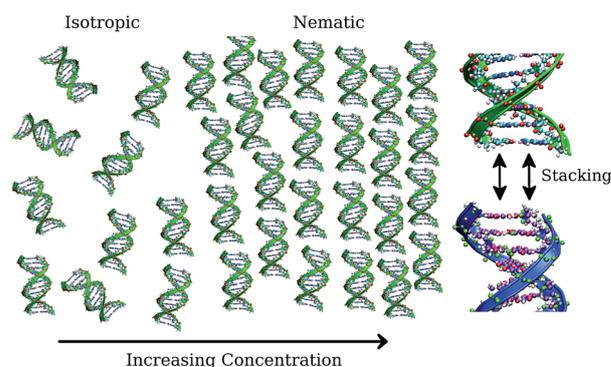
Fig. 1 Cartoon of the isotropic and nematic phases of short DNA duplexes. On increasing the density or decreasing the temperature the duplexes form persistent chains. Their growth induces the nematic transition. The polymerization is driven by the end-to-end stacking interaction between duplexes, as shown on the right side of the figure.

core molecules were shown to exhibit an unusual ferroelectric behavior. In addition, bent-core mesogens have been the first thermotropic liquid crystals for which a biaxial nematic phase (N_b) has been observed.[40,41] This suggests many possible technological applications for bent-core mesogens as functional materials.[42]

To clarify the origin of the sequence dependence of the critical concentration and the sequence dependence of the DNAD shape, we perform a numerical study of several dodecamers to evaluate their structural differences and the effect of these different structures on the isotropic–nematic (IN) transition. More specifically, we perform atomistic molecular dynamics simulations to evaluate the different degree of conformational bending as well as simulations of a coarse-grained bent-cylinder model (BC) in which the conformational bending is comparable with the values provided by the atomistic simulations. The use of a coarse-grained model makes it possible to estimate the isotropic–nematic boundaries. We complement the numerical studies with a recent theoretical approach,[43,44] properly accounting for the additional contribution to the flexibility of the chains introduced by the particle shape. The numerical and theoretical results show that the sequence dependence of the IN boundary reported in ref. 25 arises from the different degree of bending of the DNADs.

2. Numerical and theoretical methods

In this section we discuss the all-atom molecular dynamic simulations of several palindromic DNA dodecamers which are known to exhibit[25] different isotropic–nematic transition concentrations. We also present the coarse-grained model developed to describe the self-assembly process of the different DNADs and the details of the Monte Carlo (MC) methods employed to evaluate the IN boundary.

2.1 Atomistic molecular dynamics simulations

All MD simulations were carried out using the software package GROMACS[45,46] using the AMBER03 force field,[47–51] TIP3P (ref. 52) and TIP4P/2005 (ref. 53) models of water, standard parameters for ions, periodic boundary conditions and particle-mesh Ewald full electrostatics with a grid density of 0.1 nm per grid point. The van der Waals and short-range electrostatic energies were calculated using a smooth cutoff of 1 nm, and integration was performed using a 2 fs time step. For the equilibration of the system we performed simulations at fixed temperature and pressure (NPT ensemble), whilst the production runs were performed in the NVT ensemble by employing a v-rescale thermostat. In NPT simulations, the temperature was kept constant at 300 K using velocity rescaling with a stochastic term with a coupling constant of 0.1 ps. Constant pressure was maintained at 1 bar using the Berendsen algorithm. We simulated 18 identical dodecamers at a concentration of 350 mg ml^{-1} (a snapshot of an all-atom simulation is shown in Fig. 2). We repeated the calculations
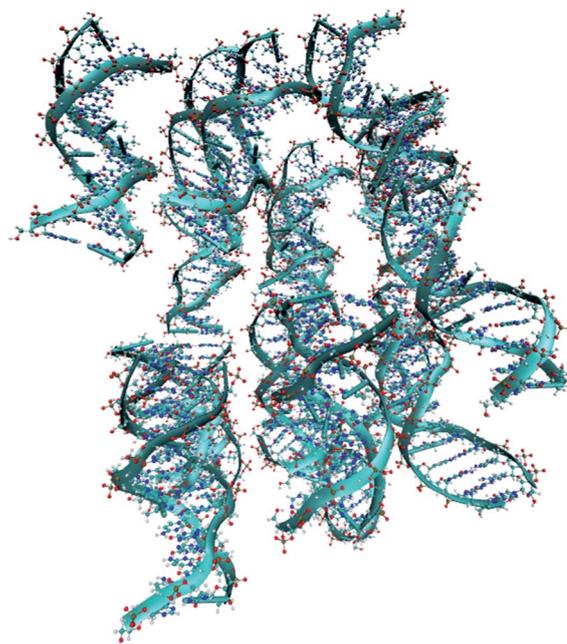
Fig. 2 Snapshot of 18 AAC dodecamers from an all-atom simulation. For the sake of clarity water molecules and ions are not shown.

Table 1 List of the DNA sequences studied by all-atom simulations, their labels and the nematic concentration at coexistence (c_N) from ref. 25

Sequence	Label	c_N (mg ml^{-1})
AATGAATTCATT	AAT	500
CGCGCCGGCGCG	allCG2	570
AATAAATTTATT	allAT	600
AACGAATTCGTT	AAC	620
CCGGCGCGCCGG	allCG1	670
CGCGAATTCGCG	DD	730
ACCGAATTCGGT	ACC	850

for seven different palindromic sequences, listed in Table 1 together with their labels and the experimentally evaluated isotropic–nematic transition concentration.[25] A graphical representation of a typical equilibrium conformation for three different sequences is reported in Fig. 3.

Na$^+$ counterions were added to neutralize the DNA charge. Na$^+$ and Cl$^-$ ions were also added in order to reach a 200 mM salt concentration. Initial double helical structures (dodecamers) have been built by the Nucleic Acid Builder which is included in the AmberTool package.[54–56] After arranging the dodecamers into the simulation box together with water molecules and ions, we performed steepest descent minimization in order to remove initial steric clashes between atoms. Then we equilibrated the system in the NPT ensemble for at least 200 ns. We check equilibration by monitoring both the total internal energy of the system and the conformation of individual dodecamers. Then we collect configurations from production runs in the canonical ensemble.
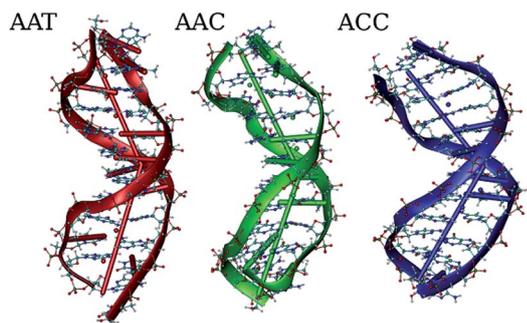
Fig. 3 Snapshot of typical equilibrium conformation extracted by all-atom MD simulations. Note the different degree of bending of the three different DNA sequences.

In order to check the dependence of our results on water models, for all sequences labelled we performed simulations with both TIP3P and TIP4P/2005 water models.

2.2 Bent cylinder models

In this work we investigated two possible coarse-grained models for DNA duplexes. The first model assimilates the dodecamer to a symmetric bent cylinder (SYBC) with diameter D, length $L_c = 2L$ (measured along the symmetry axis of each cylinder) and bending angle θ_b (expressed in degrees in the following), as shown in Fig. 4. The bent cylinder consists of two unaligned cylinders of equal length ($>L$) and equal diameter D. The length is chosen in such a way that there is no void space between the middle bases of the two cylinders. The second model assimilates DNA to an asymmetric bent cylinder (ASBC). In this second case, the bent cylinder consists of two cylinders of equal diameter but different lengths, merged together. The length measured along the two cylinder axis is $L_c = L_1 + L_2$. Also in this case, no gap is left between the two internal (middle) bases (see Fig. 4).

To model hydrophobic stacking forces between duplexes, the bent cylinders are decorated with two square-well sites. The two attractive sites are located along the symmetry axes of the two unaligned cylinders as shown in Fig. 5. The distance between an attractive site and the surface of the corresponding cylinder is $0.075D$. Two distinct sites interact *via* the square-well (SW) potential, $\beta u_{SW} = -\beta u_0$, if $r < \delta$ and $\beta u_{SW} = 0$, if $r > \delta$, where r is the distance between the interacting sites, $\delta = 0.25D$ is the interaction range (*i.e.* the diameter of the attractive sites), βu_0 is the ratio between the binding energy and the thermal energy $k_B T$ where k_B is the Boltzmann constant. The value of δ and its geometrical location control the so-called bonding volume, *i.e.* the amount of available space for bonding. In the following we will make use of the adimensional temperature $T^* = k_B T/u_0$. The pressure P is also reported as the adimensional pressure $P^* = D^3 P/8u_0$. Such values of the patch parameters (*i.e.* the attractive site position and its interaction range δ), identical to the ones chosen in a previous study,[57] enforce the single bond per site condition. By calculating the excluded volume and the persistence length of the models, it is possible to apply a recently developed theoretical framework[43] to evaluate the isotropic and nematic free energies.
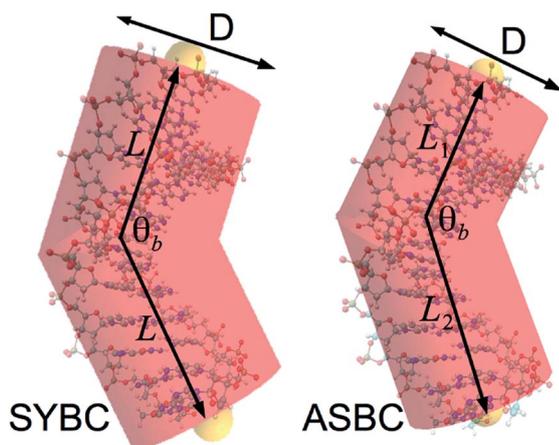
2.3 Monte Carlo simulations

We perform MC-NPT simulations and successive umbrella sampling (SUS) for the SYBC model.

2.3.1 NPT. Standard MC simulations in the NPT ensemble are performed to evaluate the equation of state for the following bending angles θ_b: 150, 155, 160, 165, 170, 175. We study 500 BCs with periodic boundary conditions. The initial configuration is prepared with all bent cylinders with the site-to-site axis (see Fig. 4) parallel to the z-axis with sizes $L_x = L_y = 10.25D$ and $L_z = 30D$, to minimize finite-size effects in the nematic phase (for a discussion on finite-size effects in a similar system see ref. 57). The NPT algorithm retained the same axis ratio as the original configuration. Simulations lasted at least 6×10^6 MC steps and thermodynamic properties were evaluated during the final 2×10^6 MC steps.

2.3.2 SUS. In order to check the phase boundaries evaluated by NPT simulations we performed a Successive Umbrella

Fig. 4 Symmetric (SYBC) and asymmetric (ASBC) bent-cylinder models of a short DNA duplex. The SYBC consists of two cylinders of equal length L whose symmetry axes form an angle θ_b. In the ASBC model, the two cylinders have different lengths L_1 and L_2 (with $L_1 + L_2 = 2L$).
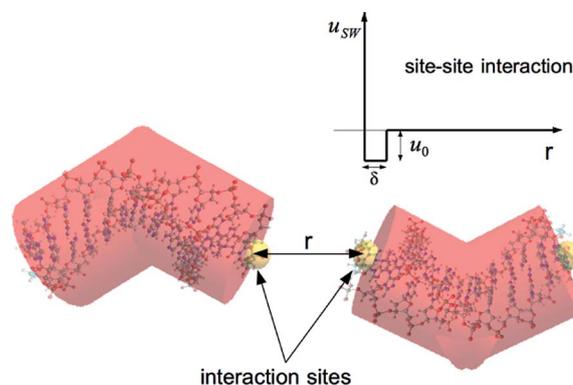
Fig. 5 Schematic representation of the interaction potential between two bent-cylinders. The square-well site–site interaction potential u_{SW}, shown in the figure, is characterized by an attractive well δ and an energy scale u_0.

Sampling[58,59] MC simulation at $T^* = 0.12$ for $\theta_b = 160$. In a SUS simulation, the probability $P(N)$ of finding N particle at fixed volume, temperature and chemical potential (*i.e.* in the grand canonical (GC) ensemble) is computed and the coexisting region of different phases can be accurately estimated.[58,59] The great benefit of SUS simulation is that $P(N)$ can be efficiently calculated through independent GC-MC simulation running in parallel by partitioning the investigated range of particle number in several overlapping windows. The whole $P(N)$ can be reconstructed matching values gathered from all the GC-MC simulations at overlapping points. $P(N)$ at different chemical potential values can be obtained by a standard histogram reweighting technique. Coexistence is defined as the condition of equal areas below the isotropic and nematic peaks.[57] The box shape in SUS simulations is not cubic, as suggested in ref. 59. Specifically we use $L_x = 25D$, $L_y = 10D$, $L_z = 50D$ where the *x*-axis is the nematic director. In the initial configuration all particles are aligned along the *x*-axis. Under these conditions, the interface builds parallel to the *xy* plane (see Fig. 10b). With this choice, chains of up to roughly 13 particles in the nematic phase do not span the box, reducing any possible finite size effect. We have checked that aggregates longer than 13 monomers do not percolate, due to the chain flexibility.

Further details about the SUS method applied to the isotropic–nematic transition can be found in ref. 57, 59 and 60.

2.4 Theory

To evaluate the phase diagram of bent-cylinder models we implement the theoretical framework which has been developed in ref. 43 and 44 and we provide here only the details which are relevant for the present discussion. More details can be found in the Appendix. According to ref. 43 and 44, the free energy of a system of equilibrium polymers[61] can be written as a sum of several contributions, namely:

$$\frac{\beta F}{V} = f_{id} + f_{exv} + f_{agg} + f_{orient} \tag{1}$$

where f_{id} is the free energy of an ideal gas of polydisperse polymers, f_{exv} is the excess contribution due to excluded volume interactions, f_{agg} models the aggregation process and f_{orient} accounts for the entropy lost in the nematic phase due to monomer alignment. The excluded volume contribution f_{exv} depends on the actual duplex conformation, *i.e.* on the bending angle θ_b; the term f_{agg} depends on the stacking free energy G_{ST}, the free energy gained on forming a bond calculated under standard thermodynamic conditions (*i.e.* $T = 293$ K and at a standard duplex concentration 1 M); f_{orient} depends on the persistence length l_p of the polymers (see eqn (17) in the Appendix). Indeed a more flexible chain will result in a larger entropy lost as monomers align in the nematic phase. Note that l_p is a function of θ_b.

In theory, the concentration of the isotropic and nematic phases at coexistence is controlled by the excluded volume and/or by the persistence length. In the following we will show that in the present case, the change of l_p on decreasing θ_b provides the dominant contribution.

3. Results and discussion

3.1 All-atom simulations

We exploit the information from all-atom simulations to estimate the best geometry of the SYBC and ASBC models. We do so for both TIP4/2005 and TIP3P water models and for seven different DNAD sequences. Results are summarized in Table 2.

In the SYBC case, we fix $L_c \approx 4$ nm (the known contour length of the DNAD with 12 bases) and estimate the angle θ_b by evaluating the end-to-end distance d_{ee} between the center of mass of the base pairs in the first and in the twelfth position along the DNA sequence. By geometry we can estimate then

$$\cos(\theta_b) = 1 - \frac{1}{2}\left(\frac{2d_{ee}}{L_{ee}}\right)^2 \tag{2}$$

where $L_{ee} = (11/12)L_c$. The use of L_{ee} instead of L_c is motivated by the following observation: the end-to-end distance d_{ee} is calculated between the two centers of mass of terminal base pairs, and hence the contour length L_{ee} associated with these two points would be approximately a base pair length, *i.e.* $(1/12)L_c$, shorter than L_c.

In the ASBC case, the three parameters requested (θ_b, L_1 and L_2) are obtained by evaluating the sum of all distances between each phosphate group in the DNA backbone with the BC surface and minimizing the resulting quantity. We constrain $L_1 + L_2$ to be within $\approx 4.0 \pm 0.02$ nm and we set the diameter of the BC to 1.86 nm. The value of L_1/L_2 for the ASBC model (last column) suggests clearly that the two unaligned cylinders can be considered of equal length, *i.e.* the DNA duplexes can be modeled as a symmetric bent cylinder (SYBC model). This result is independent of the force field used to model water, *i.e.* TIP3P or TIP4P/2005.

Table 2 Parameters of the SYBC and ASBC models evaluated from the all-atom simulations for two different water-models (TIP4P/2005 and TIP3P) for the studied DNA sequences

TIP4P/2005	SYBC	ASBC		
Sequence	θ_b	θ_b	L_1 (nm)	L_2 (nm)	L_1/L_2
AT	159.8	149.0	1.88	2.13	0.88
allCG2	144.3	143.3	1.89	2.12	0.89
allAT	151.0	146.7	1.89	2.12	0.89
AAC	157.3	148.7	1.88	2.13	0.88
allCG1	138.6	138.2	1.89	2.13	0.89
DD	145.3	144.3	1.88	2.12	0.89
ACC	130.9	132.4	1.90	2.10	0.90

TIP3P	SYBC	ASBC		
Sequence	θ_b	θ_b	L_1 (nm)	L_2 (nm)	L_1/L_2
AAT	152.8	146.3	1.89	2.12	0.89
allCG2	141.3	140.4	1.89	2.12	0.89
allAT	147.5	144.1	1.88	2.12	0.89
AAC	147.0	143.7	1.88	2.12	0.89
allCG1	146.0	142.6	1.89	2.13	0.89
DD	143.3	143.4	1.88	2.12	0.89
ACC	134.4	135.7	1.89	2.11	0.90

In Fig. 6 we plot the bending angle calculated for the dodecamers which we investigated against the IN critical concentration obtained experimentally. A clear correlation emerges between bending and critical concentration, *i.e.* the higher the bending and the lower will be the tendency of the system to nematize. The degree of bending in typical conformations for three sequences (AAT, AAC and ACC) can be seen in Fig. 3. As reported in Table 1, the concentrations of the nematic phase at coexistence c_N for AAT, AAC and ACC sequences are 500 mg ml^{-1}, 620 mg ml^{-1} and 850 mg ml^{-1} and these concentrations again correlate well with the increasing bending shown in Fig. 3 for these sequences.

3.2 Comparison between theoretical and numerical results

We performed NPT-MC simulations for the SYBC model for several bending angles at $T^* = 0.12$ and different P. The resulting equation of state (EOS) is shown in Fig. 7a. All EOS show a discontinuity signaling the phase transition between the isotropic and the nematic phase. Indeed, the nematic order parameter S (*i.e.* the largest eigenvalue of the average order tensor[62]) jumps from small ($S \approx 0.2$) to large ($S \approx 0.7$) values at the transition. The coexisting pressure, as well as the density of the two coexisting phases ϕ_I and ϕ_N increases on increasing the

Fig. 7 (a) Equation of state of the SYBC model for $\theta_b = 160$, 165 and 175 at $T^* = 0.12$ from NPT-MC simulations. The horizontal green lines indicate the approximate location of the IN transition. (b) Phase diagram of the same model as predicted by the theoretical approach (lines, see Appendix), by the NPT-MC simulations (triangles) and by the SUS calculations (circles).

particle bending, a consequence of the increased flexibility of the aggregates. Snapshots of two configurations from MC-NPT with different bending angles (shown in Fig. 8) provide clear evidence that the persistence length of the chains is highly sensitive to bending.

Fig. 6 Correlation between the bending angle θ_b and the coexisting concentration c_N of the nematic phase (from ref. 25) for ASBC (a) and SYBC (b) models. Dashed lines are a guide to the eye.
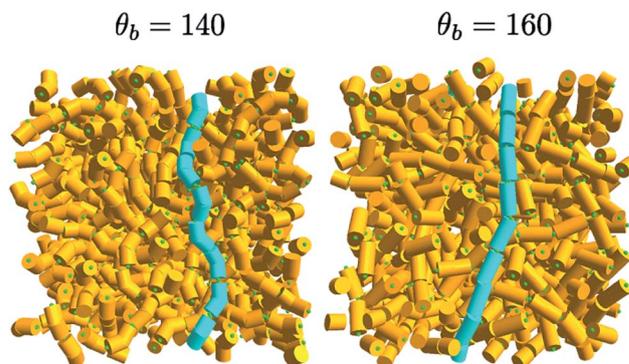
Fig. 8 Snapshots at the same volume fraction $\phi = 0.22$ and temperature $T^* = 0.12$ of the isotropic phase for the SYBC model for two different values of the bending angle ($\theta_b = 140$ and $\theta_b = 160$), to highlight the different persistence length of the selected chains (colored in cyan) in the two cases.

Fig. 7b shows the θ_b dependence of ϕ_I and ϕ_N resulting from the theoretical approach and from the NPT-MC equation of state estimates. The theory properly represents the numerical data (with a slight overestimate of ϕ_I), suggesting the possibility to extend the prediction beyond the region where numerical data are available. Indeed, the increase of ϕ_N on decreasing θ_b makes MC simulations more and more computationally demanding. The theory confirms the steep rise of the coexisting volume fractions on increasing the bending. To shed light on the physical origin of such steep rise we show in Fig. 9 the θ_b dependence of the inverse of the persistence length l_p^{-1} and compare this dependence with the θ_b dependence of the phase boundaries. Both curves show a steep rise on decreasing θ_b. To provide further evidence that indeed l_p^{-1} is the key factor in controlling this behavior we show also the theoretical phase diagram where we retain the same input parameters but fix the persistence length l_p to the value for $\theta_b = 180$. If the θ_b dependence of l_p is suppressed, the steep rise of the coexisting volume fractions disappears. This finding confirms that the significant increase in the volume fraction of the coexisting region is mostly caused by the decrease of persistence length at small bending angles.

Since NPT simulations provide an approximate coexistence boundary, being affected by thermodynamic metastability, we perform an exact evaluation of the coexisting density evaluating the density of states $P(N)$ in a grand canonical simulation. Specifically, we calculate $P(N)$ for the SYBC model by the SUS method.[58] A snapshot of a configuration in the coexistence region, displaying a clear isotropic–nematic stable interface, is shown in Fig. 10b.[59] In the presence of a stable interface, the probability $P(N)$ of observing N particles in the simulation box at fixed T and chemical potential exhibits two peaks, as shown in Fig. 10a. The logarithm of the ratio of $P(N)$ evaluated at the peak

Fig. 10 (a) Probability distribution $P(N)$, where N is the number of BCs in the simulation box, calculated from SUS simulations at $T^* = 0.12$ for $\theta_b = 160$. Here box sizes are $L_x = 25D$, $L_y = 10D$, $L_z = 50D$. (b) Snapshot of the system for $N = 2539$. Particles in the nematic phase are colored orange, while those in the isotropic phase are colored cyan.

and in the flat region between the peaks provides a measure of the surface tension.[57,60] The average over each peak of N provides an estimate of the number density (or volume fractions) of the isotropic and nematic phases at coexistence. The resulting values are also reported in Fig. 7b ϕ_N coincides with, while ϕ_I is slightly larger than the NPT estimates.

3.3 Comparison between theoretical and experimental results

To provide a theoretical prediction to be compared with experimental data we evaluate the nematic concentration at coexistence c_N for different bending angles. As discussed in the previous section, the theory requires information on the excluded volume of the particle, the persistence length and the bonding (stacking) G_{ST} free energy driving the polymerization process. Assuming bent cylinders with a diameter of 1.8 nm and a contour length of 3.6 nm (values which are compatible with the geometry of the DNA dodecamers), we can calculate the excluded volume and persistence length for several bending angles. Since there is no experimental consensus on the value of G_{ST}, we have solved the theory for two G_{ST} values, respectively $G_{ST} = -0.9$ and $G_{ST} = -2.5$ kcal mol^{-1}. The selected range of stacking energies G_{ST} is compatible with the estimate provided

Fig. 9 Dependence of the persistence length l_p on the bending angle θ_b (green line, right y-axis). The curve is superimposed on the theoretical phase diagram (full lines, the same as Fig. 7) to highlight the correlation between the volume fraction of the coexisting phases and l_p^{-1}. As a further support for the role of the persistence length in controlling the volume fraction of the coexisting phases, the figure also shows (dashed lines) the phase boundaries calculated for the same parameters as in Fig. 7 except for l_p whose value has been fixed for all bending angles to the value at $\theta_b = 180$. In this case, the volume fraction of the coexisting phases only weakly depends on θ_b.

in ref. 44. The resulting theoretical predictions for these two values are shown in Fig. 11.

To compare the theoretical predictions with the experimental results we associate with each experimentally studied dodecamer its measured c_N and a θ_b value evaluated by all-atom simulations (see Section 3.1). We report the θ_b values evaluated for both water models employed in the simulation. The resulting points fall within error inside the grey band delimiting the uncertainty in the G_{ST} values. As discussed previously, c_N increases with bending, *i.e.* with decreasing persistence length l_p of aggregates. Interestingly, this inverse scaling between c_N and l_p has been recently evidenced by experimental studies on semi-flexible amyloid fibers, which are formed by aggregation of β-lactoglobulin protein in water.[63,64]
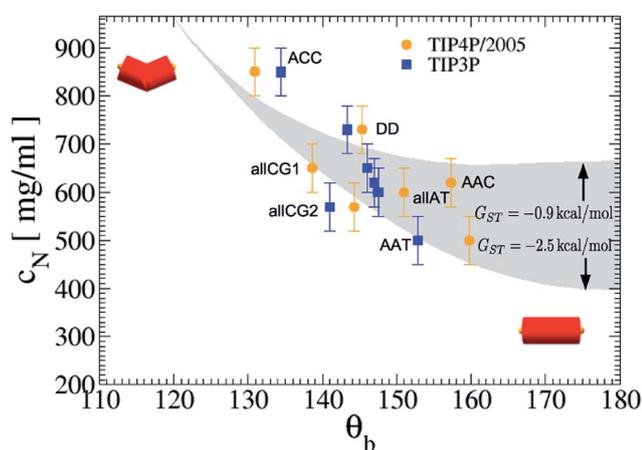
Fig. 11 Concentration of the nematic phase at coexistence c_N for all sequences studied as a function of the bending angles θ_b. The bending angle of each sequence has been estimated by all-atom simulations while the concentration c_N is the value obtained from experiments.[25] The grey band reflects the theoretical predictions for the SYBC model for $-2.5 < G_{ST} < -0.9$ kcal mol^{-1}.
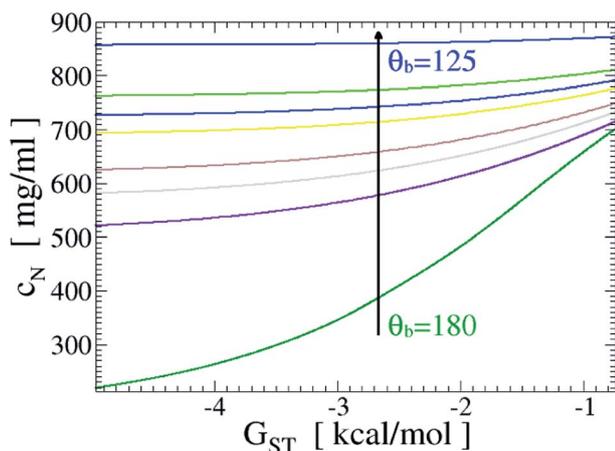
Fig. 12 Theoretical predictions for the SYBC model of the concentration of the nematic phase at coexistence c_N as a function of the stacking free-energy G_{ST} for bending angles ranging from $\theta_b = 125$ up to $\theta_b = 180$.

The theoretical predictions reported in Fig. 11 show that for low bending angles (*i.e.* $\theta_b \approx 120$) the coexisting region boundaries are rather insensitive on G_{ST}. Since G_{ST} depends on T (see eqn (5)), this prediction suggests that c_N for very bent sequences (*e.g.* ACC and allCG1) will display a very weak T dependence. The experimental observation of $c_N(T)$ could thus provide an experimental test of the theoretical framework (Fig. 12).

4. Conclusions

In this article we have studied the phase behavior of double-stranded DNA dodecamers. These DNA constructs are experimentally known to undergo an isotropic–nematic transition on increasing the concentration, due to the progressive polymerization induced by hydrophobic stacking forces acting between the exposed terminal bases.

By a combined numerical and theoretical approach we have been able to show that the experimentally observed differences between the concentrations at which the nematic transition occurs arise from the different conformational bending of the dodecamers. To do so, we have estimated the bending angle for the different sequences *via* all-atom simulations, confirming that the bending angle can differ by 20–30 degrees. The bending difference is amplified by the polymerization process, producing chains of bonded dodecamers with rather distinct persistence length. To estimate how the dodecamer bending affects the persistence length we explicitly evaluate the persistence length approximating the dodecamer with a symmetric bent cylinder model. Finally, we implement a recently proposed theoretical approach for the isotropic–nematic transition in the presence of equilibrium polymerization to evaluate the isotropic–nematic phase boundaries for the different bending angles.

We have thus been able to demonstrate that the bending of the different sequences correlates with the concentration c_N of the nematic phase at coexistence, explaining the c_N dependence experimentally observed in DNA duplexes with similar length. In addition, we show that a model in which the DNA dodecamers are represented as polymerizing symmetric bent cylinders describes the experimental results using a reasonable estimate for the base–base stacking free energy. Finally, we also show that the theoretical predictions agree rather well with the "exact" calculation of the coarse-grained bent cylinder model and hence short DNA duplexes can be viewed as bent-core mesogens. In this respect, the theoretical modeling constitutes a first attempt in the direction of developing a semi-quantitative theory for IN transition of bent-core mesogens. This opens up new perspectives in terms of possible technological applications of DNA-based liquid crystals.

Appendix A. Theory of self-assembly-driven nematization of bent-core nematogens

We build on the theoretical framework which has been developed in ref. 43 and 44 and we provide here only the details which are relevant for the present discussion. According to ref.

43 and 44, the free energy of a system of equilibrium polymers, whose various contributions have been provided in eqn (1), can be written more explicitly as follows.

$$\frac{\beta F}{V} = \sum_{l=1}^{\infty} \nu(l)\{\ln[\nu_d\nu(l)] - 1\} + \frac{\eta(\phi)}{2}\sum_{\substack{l=1 \\ l'=1}}^{\infty} \nu(l)\nu(l')\nu_{excl}(l,l')$$
$$- \beta\Delta F_b \sum_{l=1}^{\infty}(l-1)\nu(l) + \sum_{l=1}^{\infty}\nu(l)\sigma_o(l) \qquad (3)$$

where V is the volume of the system, ν_d is the volume of a monomer, $\phi \equiv \nu_d\rho$ ($\rho = N/V$ is the number density of monomers) is the packing fraction, $\nu(l)$ is the discrete number density of chains of length l, normalized such that $\sum_{l=1}^{\infty} l\nu(l) = \rho$, ΔF_b is a parameter which depends on the free energy associated with a single bond and $\nu_{excl}(l, l')$ is the excluded volume of two chains of length l and l'. $\eta(\phi)$ is the Parsons–Lee factor[65]

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

and $\sigma_o(l)$[66] accounts for the orientational entropy lost by a chain of length l in the nematic phase (including possible contribution due to its flexibility).[43,67] The parameter ΔF_b can be related to the coaxial stacking free energy G_{ST}, which can be measured experimentally, as follows:

$$G_{ST} = -\Delta F_b - k_B T \ln(\rho\nu_d). \qquad (5)$$

G_{ST} will be calculated under standard conditions, i.e. $T = 293$ K and at a standard concentration 1 M of DNADs.

A.1 Isotropic phase

In the isotropic phase, as in ref. 43, we assume the following form for the excluded volume $\nu_{excl}(l, l', X_0)$

$$\nu_{excl}(l, l', X_0) = 2B_I X_0^2 ll' + 2\nu_d k_I \frac{l+l'}{2} \qquad (6)$$

where the parameters B_I and k_I can be estimated via MC integrals of a system composed of only two monomers (see ref. 43) and X_0 is the aspect ratio of the monomers defined as follows:

$$X_0 \equiv \frac{L_{eff}}{D_{eff}} \qquad (7)$$

where L_{eff} and D_{eff} are the length and diameter of an equivalent straight cylinder having the same volume $\nu_{BC}(\theta_b)$ of the BC. Specifically the length L_{eff} is chosen to be the distance between the centers of the two attractive sites d_{ss}, i.e. $L_{eff} = d_{ss}$ and D_{eff} is such that the volume of the cylinder equals that of the BC, i.e.

$$\frac{\pi}{4}D_{eff}^2 d_{ss} = \nu_{BC}(\theta_b). \qquad (8)$$

We estimated the volume ν_{BC} via a Monte Carlo integration and the results for the BC model are shown in Fig. 13. It can be seen that ν_{BC} is substantially equal to the volume $\nu_{HC} = \nu_{BC}(180)$ of a straight HC for all bending angles.
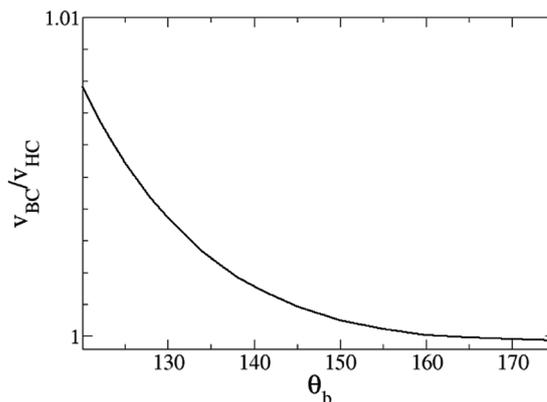
Fig. 13 BC volume as a function of bending angle θ_b.

The chain length distribution $\nu(l)$ is assumed to be exponential:[43]

$$\nu(l) = \rho M^{-(l+1)}(M-1)^{l-1} \qquad (9)$$

where the average chain length M is:

$$M = \frac{\sum_{l=1}^{\infty} l\nu(l)}{\sum_{l=1}^{\infty} \nu(l)}. \qquad (10)$$

With this choice for $\nu(l)$ the free energy in eqn (3) becomes:

$$\frac{\beta F_I}{V} = -\rho\beta\Delta F_b\left(1-M^{-1}\right)$$
$$+ \eta(\phi)\left[B_I X_0^2 + \frac{\nu_d k_I}{M}\right]\rho^2$$
$$+ \frac{\rho}{M}\left[\ln\left(\frac{\nu_d\rho}{M}\right) - 1\right]$$
$$+ \rho\frac{M-1}{M}\ln(M-1) - \rho\ln M. \qquad (11)$$

A.2 Nematic phase

In the nematic phase the monomer orientational distribution function $f(\theta)$ for the present BC model is assumed to depend only on the angle θ between the particle and the nematic axis while all orientations around such an axis are taken as equally likely. For $f(\theta)$ we use the form proposed by Onsager,[68] i.e.:

$$f_\alpha(\theta) = \frac{\alpha}{4\pi\sinh\alpha}\cosh(\alpha\cos\theta) \qquad (12)$$

where α controls the width of the angular distribution. The equilibrium value of α can be obtained by minimizing the free energy with respect to α. The excluded volume in the nematic phase takes the following form:

$$\nu_{excl}(l, l', X_0, \theta_b, \alpha) = 2B_N X_0^2 ll' + 2\nu_{BC}k_N\frac{l+l'}{2} \qquad (13)$$

where the first term and the second term on the right hand side are midsection–midsection and end–midsection contributions to the excluded volume of two BCs (see ref. 44) with

$$B_{\mathrm{N}}(\theta_{\mathrm{b}}, \alpha) = \frac{\pi}{4} D_{\mathrm{eff}}{}^3 \left(\eta_1 + \frac{\eta_2}{\alpha^{1/2}} \right) \quad (14)$$

$$k_{\mathrm{N}}(\theta_{\mathrm{b}}, \alpha) = k_{\mathrm{N}}^{\mathrm{HC}}(\alpha) + 4(\xi_1 - 1) - 4 \frac{\xi_2 - 1}{\alpha} \quad (15)$$

where $k_{\mathrm{N}}^{\mathrm{HC}}(\alpha)$ is such that $2\nu_{\mathrm{d}} k_{\mathrm{N}}^{\mathrm{HC}}(\alpha)$ is the end–midsection contribution to the excluded volume of two hard cylinders (HCs) and $\eta_k(\theta_{\mathrm{b}})$, $\xi_k(\theta_{\mathrm{b}})$ with $k = 1, 2$ are four parameters that we chose in order to reproduce the excluded volume calculated from MC calculations for all θ_{b} considered as done and discussed in ref. 43 and 44. Plugging eqn (13) and (9) into eqn (3), one obtains:

$$\frac{\beta F_{\mathrm{N}}}{V} = \hat{\sigma}_{\mathrm{o}}(l_{\mathrm{p}}) - \rho \beta \Delta F_{\mathrm{b}} \left(1 - M^{-1} \right)$$
$$+ \eta(\phi) \left[B_{\mathrm{N}}(\alpha) X_0^2 + \frac{\nu_{\mathrm{d}} k_{\mathrm{N}}^{\mathrm{HC}}(\alpha)}{M} \right] \rho^2$$
$$+ \frac{\rho}{M} \left(\ln \left[\frac{\nu_{\mathrm{d}} \rho}{M} \right] - 1 \right) - \rho \ln M$$
$$+ \rho \ln(M - 1) \frac{M - 1}{M} \quad (16)$$

where $\hat{\sigma}_{\mathrm{o}}(l_{\mathrm{p}}) \equiv \sum_l \sigma_{\mathrm{o}}(l) \nu(l)$. The orientational entropic contribution $\hat{\sigma}_{\mathrm{o}}(l_{\mathrm{p}})$ depends on the persistence length l_{p} of the chains (see ref. 43 for more details). Thus l_{p} has to be calculated for the present model for all bending angles θ_{b}. As done in ref. 43 we estimate the persistence length l_{p} by evaluating the following spatial correlation function:

$$C_{\mathrm{O}}(|i - j|) \equiv \sum_{i,j} \left\langle \hat{\mathbf{x}}(i) \cdot \hat{\mathbf{x}}(j) \right\rangle \quad (17)$$

where $\langle \ldots \rangle$ denotes an average over a set of independent random chains and i, j label two BCs along the chain ($i = 0$ is the first BC at chain end) and $\hat{\mathbf{x}}(i)$ is a unit versor parallel to direction along which the two attractive sites lie (see Fig. 4). $\langle \ldots \rangle$ denotes an average over the whole set of independent chains which has been generated. In Fig. 14 we show l_{p} as a function of θ_{b}. It can be seen that persistence length does depend significantly on θ_{b} so that one can expect that on reducing θ_{b} the IN phase boundary should shift to higher ϕ values. This decrease can be understood if one considers the following argument: consider a random chain of BCs where the bases of two successive BCs are in contact, due to BCs bending the correlation function $C_{\mathrm{O}}(|i - j|)$ goes to 0 for $|i - j| \to \infty$ but if $\theta_{\mathrm{b}} = 180$ (i.e. BC are straight cylinders) $C_{\mathrm{O}}(|i - j|)$ would remain equal to 1.

Finally we define the usual nematic order parameter S which is related to α as follows:

$$S(\alpha) = \int (3 \cos^2 \theta - 1) f_\alpha(\theta) \pi \sin \theta \mathrm{d}\theta \approx 1 - 3/\alpha. \quad (18)$$

A.3 Phase coexistence

Phase boundaries of IN transition are characterized by coexisting isotropic and nematic phases in which the volume
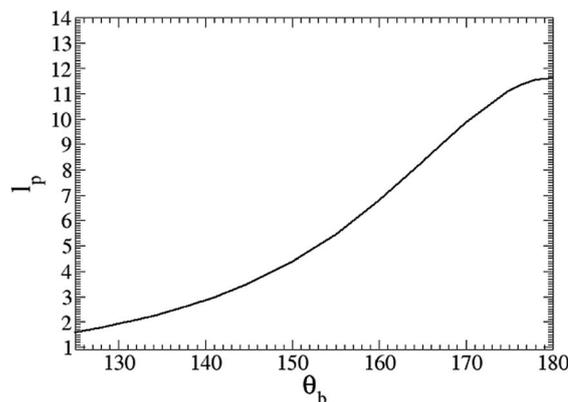
Fig. 14 Persistence length l_{p} (in unit of monomer) as a function of bending angle θ_{b}.

fractions of BCs are, respectively, $\phi_{\mathrm{N}} = \nu_{\mathrm{BC}} \rho_{\mathrm{N}}$ and $\phi_{\mathrm{I}} = \nu_{\mathrm{BC}} \rho_{\mathrm{I}}$. ρ_{I} and ρ_{N} can be calculated by the following set of equations:

$$\frac{\partial}{\partial M_{\mathrm{I}}} F_{\mathrm{I}}(\rho_{\mathrm{I}}, M_{\mathrm{I}}) = 0$$
$$\frac{\partial}{\partial M_{\mathrm{N}}} F_{\mathrm{N}}(\rho_{\mathrm{N}}, M_{\mathrm{N}}, \alpha) = 0$$
$$\frac{\partial}{\partial \alpha} F_{\mathrm{N}}(\rho_{\mathrm{N}}, M_{\mathrm{N}}, \alpha) = 0 \quad (19)$$
$$P_{\mathrm{I}}(\rho_{\mathrm{I}}, M_{\mathrm{I}}) = P_{\mathrm{N}}(\rho_{\mathrm{N}}, M_{\mathrm{N}}, \alpha)$$
$$\mu_{\mathrm{I}}(\rho_{\mathrm{I}}, M_{\mathrm{I}}) = \mu_{\mathrm{N}}(\rho_{\mathrm{N}}, M_{\mathrm{N}}, \alpha).$$

The first three equations express the fact that eqn (11) has to be minimized with respect to M_{I} and eqn (16) with respect to M_{N} and α. The remaining two equations impose the equal pressure and chemical potential conditions for the two phases at equilibrium, i.e. $P_{\mathrm{I}} = P_{\mathrm{N}}$ and $\mu_{\mathrm{I}} = \mu_{\mathrm{N}}$.

Acknowledgements

References

1 N. C. Seeman, *J. Theor. Biol.*, 1982, **99**, 237–247.
2 J. H. Chen and N. C. Seeman, *Nature*, 1991, **350**, 631–633.
3 K. VanWorkum and J. F. Douglas, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **73**, 031502.
4 C. Mirkin, R. Letsinger, R. Mucic and J. Storhoff, *Nature*, 1996, **382**, 607.
5 V. N. Manoharan, M. T. Elsesser and D. J. Pine, *Science*, 2003, **301**, 483–487.

6 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.

7 G. Yi, V. N. Manoharan, E. Michel, M. T. Elsesser, S. Yang and D. J. Pine, *Adv. Mater.*, 2004, **16**, 1204–1207.

8 F. W. Starr, J. F. Douglas and S. C. Glotzer, *J. Chem. Phys.*, 2003, **119**, 1777–1788.

9 F. W. Starr and F. Sciortino, *J. Phys.: Condens. Matter*, 2006, **18**, L347–L353.

10 S. I. Stupp, S. Son, H. C. Lin and L. S. Li, *Science*, 1993, **259**, 59–63.

11 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.

12 P. W. K. Rothemund, *Nature*, 2006, **440**, 297–302.

13 N. C. Seeman, *Nature*, 2003, **421**, 427–431.

14 A. V. Tkachenko, *Phys. Rev. Lett.*, 2002, **89**, 148303.

15 M. R. Jones, R. J. Macfarlane, B. Lee, J. Zhang, K. L. Young, A. J. Senesi and C. A. Mirkin, *Nat. Mater.*, 2010, **9**, 913–917.

16 S. Y. Park, A. K. R. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz and C. A. Mirkin, *Nature*, 2008, **451**, 553–556.

17 D. Nykypanchuk, M. M. Maye, D. van der Lelie and O. Gang, *Nature*, 2008, **451**, 549–552.

18 S. Sacanna, W. T. M. Irvine, P. M. Chaikin and D. J. Pine, *Nature*, 2010, **464**, 575–578.

19 T. Bellini, R. Cerbino and G. Zanchetta, *DNA-based Soft Phases*, Springer, Berlin, Heidelberg, 2011, pp. 1–55.

20 S. H. Um, J. B. Lee, N. Park, S. Y. Kwon, C. C. Umbach and D. Luo, *Nat. Mater.*, 2006, **5**, 797–801.

21 L. Rovigatti, F. Smallenburg, F. Romano and F. Sciortino, *ACS Nano*, 2014, **8**, 3567–3574.

22 S. Biffi, R. Cerbino, F. Bomboi, E. M. Paraboschi, R. Asselta, F. Sciortino and T. Bellini, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 15633–15637.

23 M. Nakata, G. Zanchetta, B. D. Chapman, C. D. Jones, J. O. Cross, R. Pindak, T. Bellini and N. A. Clark, *Science*, 2007, **318**, 1276–1279.

24 G. Zanchetta, M. Nakata, M. Buscaglia, N. A. Clark and T. Bellini, *J. Phys.: Condens. Matter*, 2008, **20**, 494214.

25 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.

26 N. F. Bouxsein, C. Leal, C. S. McAllister, K. K. Ewert, Y. Li, C. E. Samuel and C. R. Safinya, *J. Am. Chem. Soc.*, 2011, **133**, 7585–7595.

27 G. Zanchetta, T. Bellini, M. Nakata and N. A. Clark, *J. Am. Chem. Soc.*, 2008, **130**, 12864–12865.

28 M. A. Young, G. Ravishanker and D. L. Beveridge, *Biophys. J.*, 1997, **73**, 2313–2336.

29 A. A. Gorin, V. B. Zhurkin and K. Wilma, *J. Mol. Biol.*, 1995, **247**, 34–48.

30 T. E. Haran, J. D. Kahn and D. M. Crothers, *J. Mol. Biol.*, 1994, **244**, 135–143.

31 M. B. Ros, J. L. Serrano, M. R. de la Fuente and C. L. Folcia, *J. Mater. Chem.*, 2005, **15**, 5093–5098.

32 Y. Lansac, P. K. Maiti, N. A. Clark and M. A. Glaser, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2003, **67**, 011703.

33 A. Dewar and P. J. Camp, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2004, **70**, 011704.

34 R. Berardi, L. Muccioli, S. Orlandi, M. Ricci and C. Zannoni, *J. Phys.: Condens. Matter*, 2008, **20**, 463101.

35 P. J. Camp, M. P. Allen and A. J. Masters, *J. Chem. Phys.*, 1999, **111**, 9871–9881.

36 H. Takezoe and Y. Takanishi, *Jpn. J. Appl. Phys.*, 2006, **45**, 597–625.

37 J. Matraszek, J. Mieczkowski, J. Szydlowska and E. Gorecka, *Liq. Cryst.*, 2000, **27**, 429–436.

38 N. Gimeno, A. Sánchez-Ferrer, N. Sebastián, R. Mezzenga and M. B. Ros, *Macromolecules*, 2011, **44**, 9586–9594.

39 T. Niori, T. Sekine, J. Watanabe, T. Furukawa and H. Takezoe, *J. Mater. Chem.*, 1996, **6**, 1231–1233.

40 B. R. Acharya, A. Primak and S. Kumar, *Phys. Rev. Lett.*, 2004, **92**, 145506.

41 C. Tschierske and D. J. Photinos, *J. Mater. Chem.*, 2010, **20**, 4263–4294.

42 J. Etxebarria and M. Blanca Ros, *J. Mater. Chem.*, 2008, **18**, 2919–2926.

43 C. De Michele, T. Bellini and F. Sciortino, *Macromolecules*, 2012, **45**, 1090–1106.

44 C. De Michele, L. Rovigatti, T. Bellini and F. Sciortino, *Soft Matter*, 2012, **8**, 8388–8398.

45 H. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.

46 S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.

47 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.

48 E. J. Sorin and V. S. Pande, *Biophys. J.*, 2005, **88**, 2472–2493.

49 A. J. DePaul, E. J. Thompson, S. S. Patel, K. Haldeman and E. J. Sorin, *Nucleic Acids Res.*, 2010, **38**, 4856–4867.

50 T. E. Cheatham and D. A. Case, *Biopolymers*, 2013, **99**, 969–977.

51 J. W. Ponder and D. A. Case, *Protein Simulations*, Academic Press, 2003, vol. 66, pp. 27–85.

52 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

53 J. L. F. Abascal and C. Vega, *J. Chem. Phys.*, 2005, **123**, 234505.

54 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **3**, 198–210.

55 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.

56 D. Case, V. Babin, J. Berryman, R. Betz, Q. Cai, D. Cerutti, T. C. III, T. Darden, R. Duke, H. Gohlke, A. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. Merz, F. Paesani, D. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. Simmerling, W. Smith, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu

and P. Kollman, *AMBER 14*, University of California, San Francisco, 2014.

57 K. T. Nguyen, F. Sciortino and C. De Michele, *Langmuir*, 2014, **30**, 4814–4819.

58 P. Virnau and M. Muller, *J. Chem. Phys.*, 2004, **120**, 10925–10930.

59 R. L. C. Vink and T. Schilling, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2005, **71**, 051716.

60 R. L. C. Vink, S. Wolfsheimer and T. Schilling, *J. Chem. Phys.*, 2005, **123**, 074901-1–074901-9.

61 F. Sciortino, J. Douglas and C. De Michele, *J. Phys.: Condens. Matter*, 2008, **20**, 155101.

62 T. Kouriabova, M. Betterton and M. Glaser, *J. Mater. Chem.*, 2010, **20**, 10366–10383.

63 R. Mezzenga, J.-M. Jung and J. Adamcik, *Langmuir*, 2010, **26**, 10401–10405.

64 J.-M. Jung and R. Mezzenga, *Langmuir*, 2010, **26**, 504–514.

65 J. D. Parsons, *Phys. Rev. A*, 1979, **19**, 1225–1230.

66 T. Odijk, *Macromolecules*, 1986, **19**, 2313–2329.

67 A. Khokhlov and A. Semenov, *Phys. A*, 1981, **108A**, 546–556.

68 L. Onsager, *Ann. N. Y. Acad. Sci.*, 1949, 627.