Peer-reviewed version available at Polymers 2018, 10, 447; doi:10.3390/polym10040447

Article

An accurate estimate of the free energy and phase diagram of all-DNA bulk fluids

Emanuele Locatelli [1,*] and Lorenzo Rovigatti [2]

[1] Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria; emanuele.locatelli@univie.ac.at
[2] CNR-ISC, Uos Sapienza, Piazzale A. Moro 2, 00185 Roma, Italy and Department of Physics, Sapienza, Università di Roma.; lorenzo.rovigatti@uniroma1.it
* Correspondence: emanuele.locatelli@univie.ac.at;

Abstract: We present a numerical study in which large-scale bulk simulations of self-assembled DNA constructs have been carried out with a realistic coarse-grained model. The investigation aims at obtaining a precise, albeit numerically demanding, estimate of the free energy for such systems. We then, in turn, use these accurate results to validate a recently proposed theoretical approach that builds on a liquid-state theory, the Wertheim theory, to compute the phase diagram of all-DNA fluids. This hybrid theoretical/numerical approach, based on the lowest order virial expansion and a nearest-neighbor DNA model, can provide, in an undemanding way, a thermodynamic description of DNA associating fluids that is in semi-quantitative agreement with experiments. We show that the predictions of such scheme are as accurate as the ones obtained with more sophisticated methods. We also demonstrate the flexibility of the approach by incorporating non-trivial additional contributions that go beyond the nearest-neighbor model to compute the DNA hybridization free energy.

Keywords: DNA; DNA Nanotechnology; patchy particles; Wertheim theory; thermodynamic integration; phase coexistence;

1. Introduction

The most prominent function of DNA is to store the genetic information of all living organisms. However, the main biophysical features that allow DNA to fulfil its role, namely the extremely regular structure of its double-stranded form and the outstanding specificity of its single-stranded form, make it a great addition to the toolboxes of nanotechnology and materials science [1,2]. The usage of DNA in the latter field is extremely variegated, since it can be used not only as a link to connect nano- or micro-particles [3,4], but also to synthesise all-DNA materials. In both cases, the resulting materials can be made either ordered or disordered [5–8].

The strategy devised to synthesise all-DNA materials is based on the concept of hierarchical or multi-step self-assembly: short (usually 10 – 100 nucleotides) DNA single strands are designed to self-assemble into well-defined object at intermediate temperature T. As T is lowered, these DNA constructs start to aggregate and form higher-order finite structures or even bulk materials [1]. Since the melting T of DNA is highly dependent on the specific strand sequence and its length, the T at which the self-assembly of the constructs and the formation of inter-construct bonds occur can be tuned independently. Therefore, it is common to use sequences of length 4–10 for the inter-construct bonding and of length 15–30 for the constructs themselves. Sequences designed to follow this principle have been used to synthesise crystals [7], gels [8] and reentrant gels [9]. Most of the times these synthetic sequences have to be chosen carefully and through a tedious and expensive trial-and-error procedure. Indeed, from a theoretical standpoint, there are two-state thermodynamic descriptions of DNA, the

so-called nearest-neighbour models, that make it easy to estimate the melting temperature (and even secondary structure) of DNA systems [10]. However, there is no straightforward way of predicting the actual phase behaviour of complicated DNA systems. For all-DNA constructs that interact and bind through a small number of short sticky ends, some progress has been made by employing a well-known liquid-state theory, the Wertheim thermodynamic perturbation theory (WTPT) [11,12], complemented by detailed simulations of a realistic DNA model [13]. It has been shown that the hybrid theoretical/numerical approach is able to semi-quantitatively (and, in some cases, quantitatively) reproduce experimental results. However, some of the assumptions and approximations upon which the method developed in Ref. [13] rests are not fully controlled and require a more stringent validation in light of their sheer simplicity. In this paper we employ large-scale molecular dynamics simulations to complement the WTPT and compute the phase behaviour of all-DNA systems with higher accuracy, in order to more strictly validate the method introduced in Ref. [13]. We also show that the method can be readily extended by considering additional contributions. Here we show that a closer agreement between theory and experiment can be obtained by taking into account the effect of the presence of the nanostars on the hybridisation of the sticky ends [14].

The paper is organized as follows: in section 2, after discussing the estimation of the Helmoltz free energy within the Wertheim TPT framework, we briefly discuss the determination of critical points and phase coexistence; we futher briefly review the numerical models. In section 3 we detail all the results obtained. First, in section we characterize the structure of the fluid (Section 3.1); we then examine and comment the data obtained from numerical simulation for the pressure as function of the concentration at different salt concentrations and different temperatures (Section 3.2). Last, we compute the critical points for trivalent and tetravalent particles, comparing the results with previous estimates (Section 3.3).

2. Methods

2.1. DNA sequences

We consider systems made up of all-DNA nanostars, that is, branched polymers with f arms. These DNA constructs are composed of f single strands with specifically designed sequences (see Table 1). The sequences used are the same reported in previous experimental works [8,15,16].

We focus on trivalent (trimers, $f = 3$) and tetravalent (tetramers, $f = 4$) constructs. In both cases each of the f strands is composed of 49 nucleotides. Each sequence is divided into two central regions (20 nucleotides each), and a third, shorter region (6 nucleotides). The two central regions are designed to form the stiff double-stranded sections of the arms; they are separated by two unpaired nucleotides, also called spacers, that provide flexibility at the centre of the construct. The third region is identical in all four sequences; one spacer is present before the sticky end, providing as well flexibility. The final sequence is self-complementary and thus allows for inter-tetramer bonding. The difference in length between the double-stranded arms and the sticky ends provides a separation between the temperatures at which bonds and arms melt. This allows for a temperature-controlled hierarchical self-assembly scenario, where the f single strands can spontaneously assemble at intermediate temperatures and form stable constructs before bonds can be formed. Therefore, there is a range of temperatures in which the construct is stable and can represent an all-DNA experimentally accessible realization of a patchy particle[8,9,17]. As will be also mentioned later, in this work we scramble the sticky end sequences so that no bonding is possible. Such expedient provides, in the case of DNA constructs, the so-called *reference fluid*.

2.2. Numerical methods

We perform simulations of DNA nanostars modelled with oxDNA2 [18], a DNA model coarse-grained at the level of single nucleotides. The interaction forms and parameters in oxDNA2 are chosen to reproduce structural and thermodynamical properties of both single- and double-stranded

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 March 2018 doi:10.20944/preprints201803.0203.v1

Peer-reviewed version available at Polymers 2018, 10, 447; doi:10.3390/polym10040447

Trimer sequences
CTACTATGGCGGGTGATAAAAACGGGAAGAGCATGCCCATCCACGATCG
GGATGGGCATGCTCTTCCCGAACTCAACTGCCTGGTGATACGACGATCG
CGTATCACCAGGCAGTTGAGAATTTATCACCCGCCATAGTAGACGATCG
Tetramer sequences
CTACTATGGCGGGTGATAAAAACGGGAAGAGCATGCCCATCCACGATCG
GGATGGGCATGCTCTTCCCGAACTCAACTGCCTGGTGATACGACGATCG
CGTATCACCAGGCAGTTGAGAACATGCGAGGGTCCAATACCGACGATCG
CGGTATTGGACCCTCGCATGAATTTATCACCCGCCATAGTAGACGATCG

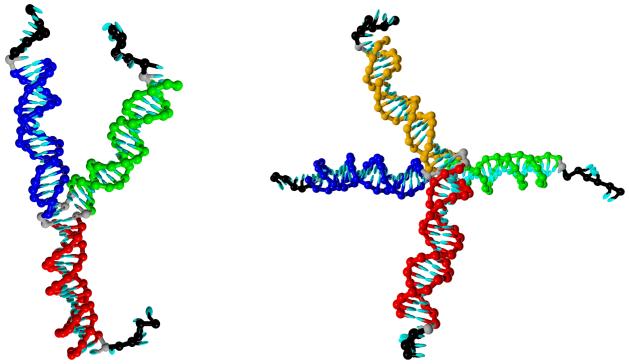

Figure 1. (Top) strand sequences for all the DNA construct types used in this work. Spacers (nucleotide that add flexibility to the structure but are not designed to pair up) are coloured in grey. Non-grey nucleotides that are part of complementary sequences share the same colour. (Bottom) Simulation snapshots of a trimer and a tetramer, coloured according to the table above.

DNA molecules in B-form. The interactions between nucleotides, modelled as rigid bodies, account for excluded volume, electrostatic repulsion between the negatively charged backbones, 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 on the structure and thermodynamics of DNA. [19,20] The electrostatic repulsion is provided by a Yukawa term characterised by a screening length which is an increasing function of T and a decreasing function of the salt concentration S [20].

We run large-scale simulations (250 purely-repulsive trimers or tetramers for a total of 36750 or 49000 nucleotides, respectively) at eight different densities, three salt concentrations and three temperatures. We perform molecular dynamics simulations in the NVT ensamble, using a Brownian thermostat [21]. We performed these simulations on NVIDIA GTX1080 GPUs [22]. In order to calculate the equation of state of the systems with adequate precision we run for $\sim 10^9 - 10^{10}$ MD steps, which took 1-2 weeks of computer time. The pressure have been computed by using the *molecular* definition of the virial [23].

2.3. Theoretical framework

We combine Wertheim TPT with an accurate mass-action law describing DNA binding. In a pure, one-species system, within the representation of the Helmoltz free energy, the thermodynamic quantities can be expressed as functions of the temperature T and the density ρ. In the WTPT the Helmholtz free energy per particle for DNA nanostars of valence $f = 4$ (tetramers) or $f = 3$ (trimers) can be written as [13,24]

$$\beta\bar{f}(T,\rho) \equiv \beta\frac{F(T,\rho)}{N} = \beta f_{\rm ref}(T,\rho) + \beta f_b(T,\rho) \tag{1}$$

where $\beta = 1/k_BT$, k_B is the Boltzmann constant. Following the standard notation, $\beta f_{\rm ref}(T,\rho)$ is the free-energy per particle of the reference system, *i.e.* the system in which particles are not decorated

Peer-reviewed version available at Polymers 2018, 10, 447; doi:10.3390/polym10040447

with bonding sites, and $\beta f_b(T,\rho)$ is the free energy per particle associated to the bond formation. The former term, in the case of DNA nanostars, is evaluated by considering a system in which the sticky end sequences are scrambled in such a way that Watson-Crick pairing does not occur and no inter-star bond can form; the residual interaction is purely repulsive. $\beta f_{\rm ref}$ itself is given by the two contributions

$$\beta f_{\rm ref}(T,\rho) = \beta f_{\rm id}(T,\rho) + \beta f_{\rm ex}(T,\rho), \tag{2}$$

where the ideal gas free-energy density is given by $\beta f_{\rm id}(T,\rho) = \ln(v_0\rho) - 1$, where v_0 is a reference volume whose value has no effect on the derivatives of the free energy. Since WPTP is often applied to systems with hard-core excluded volume interactions, the free energy of the reference fluid is often approximated with the Carnahan-Starling expression. Here, however, we deal with a system for which there are known theoretical estimates. In Ref. [13] a simple virial expansion truncated at the second order was adopted for the sake of simplicity and of numerical efficiency. Here we extensively check the quality of the approximation by providing more accurate estimates of the reference free energy. More specifically, we compute the excess free energy per particle following two standard routes: thermodynamic integration (TI) and the virial expansion truncated at the third order.

The standard path of thermodynamic integration consists in computing the free energy from a known reference state through integration of the equation of state as

$$\beta f_{\rm ref}(T,\rho) = \beta f_{\rm id}(T,\rho) + \beta f_0(T,\rho) + \int_0^\rho \frac{\beta P(\rho') - \rho'}{\rho'^2} d\rho' \tag{3}$$

We compute the pressure P in bulk numerical simulations at different densities ρ, fixed temperature and salt concentration; we then interpolate and integrate the pressure analytically within the range of densities considered. The pressure measurements done in the bulk, large-scale numerical simulation performed can be also used to improve the virial expansion, which was previously limited at the second order [13]. Within the virial expansion, the pressure as function of density reads

$$\beta P = \rho + B_2\rho^2 + B_3\rho^3 \tag{4}$$

where $B_2 = B_2(T)$ is the second virial coefficient

$$B_2(T) = -\frac{1}{2}\int_0^\infty 4\pi r^2 \left(\exp\left(-\frac{V(r)}{k_B T}\right) - 1\right) dr \tag{5}$$

computed at different salt concentrations; $V(r)$ is here the effective intra-molecular pair potential. We obtain the third virial coefficient through a fitting procedure; then we write the excess free energy as in Ref. [13], employing the virial expansion

$$\beta f_{\rm ex}(T,\rho) = B_2(T)\rho + \frac{B_3(T)}{2}\rho^2 \tag{6}$$

Finally, as all bonds are identical, the bonding free-energy per particle is given by [25]

$$\beta f_b(T,\rho) = f\left(\ln(1 - p_b(T,\rho)) + \frac{1}{2}p_b(T,\rho)\right) \tag{7}$$

where p_b is the fraction of formed bonds. The latter, which is function of T and ρ, can be evaluated *via* a law of mass action, yielding

$$p_b(T,\rho) = 1 - \frac{-1 + \sqrt{1 + 4\Delta f\rho}}{2\Delta f\rho} \tag{8}$$

where Δ is linked to the free-energy difference between bonded and non-bonded pairs of sticky ends. Following Ref. [10,26]

$$\Delta \equiv v_b \exp(-\beta \Delta G) = v_b \exp\left(-\frac{\Delta H - (\Delta S_{\text{nosalt}} + \Delta S_{\text{salt}})T}{k_B T}\right) \tag{9}$$

where $v_b = 1.6606\,\text{nm}^3$ is the reference volume of the non-bonded single strands [27], $\Delta H = -54000$ cal is the enthalpy gain upon bonding and $\Delta S_{\text{nosalt}} = -151.99$ cal/K and $\Delta S_{\text{salt}} = 0.368 \cdot (L_{\text{DNA}} - 1) \cdot \ln(S)$ cal/K are the salt-independent and salt-dependent entropy variations again upon bonding, respectively. These quantities refer to the sticky end sequences considered here, for which $L_{\text{DNA}} = 6$ (see Appendix for additional details on the different contributions to the bonding free energy). The expression for Δ thus encodes the salt, temperature and sequence dependence of the free-energy difference between bonded and non-bonded states. In general, the free energy computed according to the SantaLucia model does not take into account the effect of the presence of more than one non-binding nucleotides flanking the binding sequence [10]. Here the sticky ends are attached to the whole construct which, especially at low salt concentrations, might exert a substantial electrostatic repulsion that it is not straightforward to quantify in terms of contributions to the binding free energy. However, a recent joint numerical/experimental work provided a quantitative estimation of the effect of single-stranded tails on the melting temperature and hybridisation free energy of short strands. [14] It was shown that the effect depends strongly on salt concentration and seems to saturate when the length of the tail exceeds 4–5 single-stranded nucleotides. Unfortunately, the effect of double-stranded tails has not been investigated yet. In order to tentatively estimate the effect of the nanostar's electrostatic repulsion on the phase behaviour we use the long-tails limit of the values reported in Ref. [14] and extrapolate its dependence on S to correct the SantaLucia's free energies. The detailed derivation of these additional terms is reported in the Appendix.

Once the free energy is known, the location of the critical point at fixed salt concentration is found invoking global stability; mathematically, the critical temperature T_c and density ρ_c must satisfy the conditions

$$\left.\frac{\partial P}{\partial \rho}\right|_{(T_c,\rho_c)} = 0 \qquad \text{and} \qquad \left.\frac{\partial^2 P}{\partial \rho^2}\right|_{(T_c,\rho_c)} = 0 \tag{10}$$

where $P(\rho, T) = \rho^2 \frac{\partial \bar{f}}{\partial \rho}$. We also evaluate the coexistence region by employing a standard common tangent construction. [24] Different points that yield phases with the same pressure and chemical potentials lie on a *tie line*. We compute the pressure and chemical potential as

$$P = -\frac{\partial \bar{f}}{\partial v} = \rho^2 \frac{\partial \bar{f}}{\partial \rho} \qquad \text{and} \qquad \mu = g = \bar{f} + \frac{P}{\rho}$$

where g is the Gibbs free energy per particle, and we exploit the well known property $\mu = g$. We thus impose

$$P_1(T, \rho_1) = P_2(T, \rho_2) \qquad \text{and} \qquad \mu_1(T, \rho_1) = \mu_2(T, \rho_2) \tag{11}$$

where P_j and μ_j are the pressure and chemical potentials in the phase j, respectively. In order to solve the system of equations given by Eq. (10) and (11), we use a nonlinear solver (modified Powell algorithm) and set a grid of initial conditions in the ρ-T or ρ_1-ρ_2 plane at fixed T, for critical and coexistence points, respectively. Such a research strategy is proficient in finding critical and coexistence points.

3. Results

3.1. Structural properties of the fluid

We consider the reference fluid as a polymeric system composed of purely repulsive particles.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 March 2018 doi:10.20944/preprints201803.0203.v1

Peer-reviewed version available at Polymers 2018, 10, 447; doi:10.3390/polym10040447

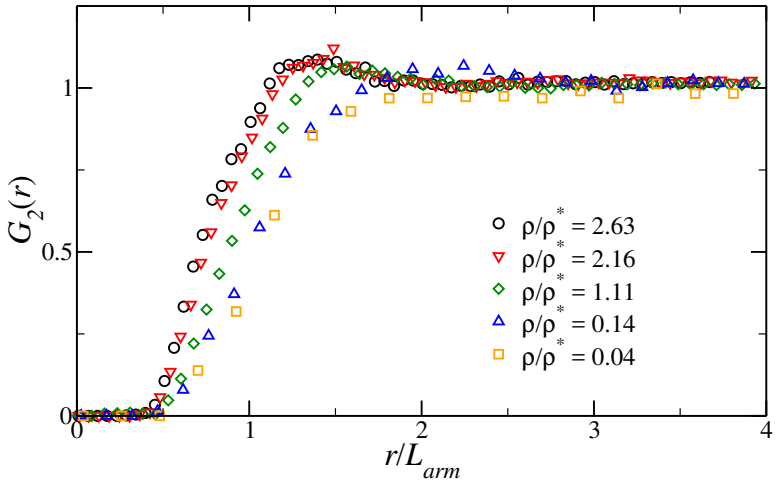

Figure 2. Radial distribution function of the centers of mass of the DNA tetramers as function of the inter-particle distance r, measured in units of the average arm length $L_{\rm arm}$, at fixed temperature $T =$ 25 °C and salt concentration $S = 0.5$M, for different reduced densities ρ/ρ^*.

We choose as the characteristic length scale of the fluid the length of one arm, defined as the average distance between the central nucleotides and any arm tip. This choice is dictated by the fact that the arms of the stars are quite rigid, especially at low salt concentration, as their length ($\approx$ 20 nucleotides) is only a fraction of the dsDNA persistence length ($\approx$ 150). On the contrary, the overall structure, due to the presence of the central unpaired nucleotides, can be very floppy [24]; thus the radius of gyration underestimates the size of the construct, and does not provide a reliable measure for the typical interaction range. From numerical simulations we find that $L_{\rm arm}$ depends on the salt concentration only, since we obtain $L_{\rm arm} = 9.70$ nm for $S = 0.05$ M, $L_{\rm arm} = 8.93$ nm for $S = 0.2$ M, $L_{\rm arm} = 8.72$ nm for $S = 0.5$ M, regardless of temperature and functionality. It follows that the overlap density for our system, defined as $\rho^* = 3/(4\pi L_{\rm arm}^3)$, is $\rho^* = 2.61 \cdot 10^{-4}$ nm^{-3} for $S = 0.5$ M, $\rho^* = 3.35 \cdot 10^{-4}$ nm^{-3} for $S = 0.2$ M and $\rho^* = 3.60 \cdot 10^{-4}$ nm^{-3} for $S = 0.5$ M.

We mention that, for the most dense systems considered, the arms become shorter and, overall, the nanostars start shrinking. This is usually a sign of the onset of the semi-dilute regime. Indeed, examining the radial distribution function $G_2(r)$, we note a structural change as $\rho/\rho^* \gtrsim 1$. As visible in Fig. 2, for very low values of ρ/ρ^*, i.e. in the dilute regime, the radial distribution function has no peak, the fluid having no structure. Upon approaching the overlap density the fluid becomes more structured and a small peak appears. This peak happens at around $1.5L_{\rm arm}$, as commonly observed in soft fluids at these densities [28]. These features supports the choice of $L_{\rm arm}$ as the characteristic of a nanostar in solution.

3.2. Pressure of the reference fluid

Next we analyse the pressure computed from large-scale numerical simulations introduced in Section 2. As reported in the literature [29], the reduced (osmotic) pressure of semi-dilute solutions of polyelectrolites with added salt should be a universal function of the reduced density $B_2\rho$, B_2 being the second virial coefficient.

In Fig. 3, we report the reduced pressure $\beta P/\rho$ as function of the reduced density $B_2\rho$. We observe a nice collapse for both tetramers and trimers, at three different salt concentrations and three different temperatures. In what follows we will use this data to compute the reference free energy, using the two methods described in Section 2.

We start from the third-order virial approximation: panel a) of Fig. 4 shows the comparison between numerical data for the reduced pressure and the best fit of Eq. (4) as function of the reduced

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 March 2018 doi:10.20944/preprints201803.0203.v1

Peer-reviewed version available at Polymers 2018, 10, 447; doi:10.3390/polym10040447

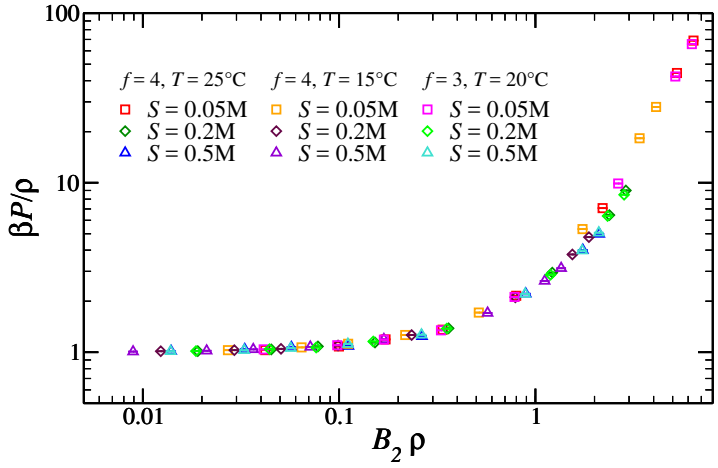

Figure 3. Reduced pressure for the reference fluid as function of the reduced density $B_2\rho$, for different functionalities, salt concentrations and temperatures.

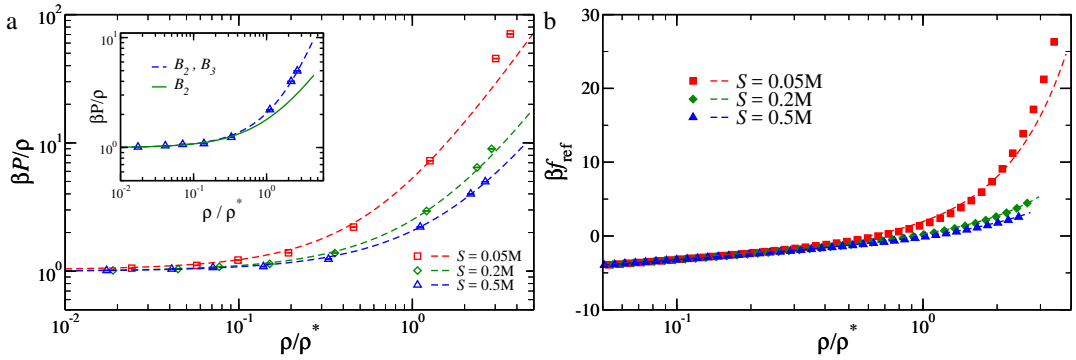

Figure 4. a) Reduced pressure of a reference fluid of tetramers, as function of the reduced density ρ/ρ^*. Points are data from numerical simulations, dashed lines are polynomial fits (Eq. (4)). Inset: Pressure as function of ρ/ρ^* for $S = 0.5$ M: symbols and dashed line as in the main panel, the full line is the second order virial approximation. b) Reference free energy as function of the reduced density ρ/ρ^* computed using TI (full symbols) and a third order virial approximation (dashed lines).

density. The dashed lines in Fig. 4a refer to polynomial fits of Eq. (4). In the fitting procedure, we impose $k_bT = T_{abs}/3000$, with T_{abs} being the absolute temperature (measured in K), as set in oxDNA; furthermore, we fix B_2 to the values reported in Ref. [13]. The only fitting parameter is thus the third virial coefficient B_3: as shown in Fig. 3, such functional form fits the numerical data well for all the densities and salt concentrations considered; a small discrepancy is observed only for the lowest salt concentration considered, well in the semi-dilute regime. We further show, in the inset of Fig. 4a, how the second order virial approximation performs: the solid line is Eq. (4), considered only up to the second order. The second order approximation fits the data very nicely almost up to the semi-dilute regime, $\rho/\rho^* = 1$, where the effective pair potential description for polymer solutions usually breaks down[30,31].

We further carry on the validation by comparing how the third-order virial approximation compares against thermodynamic integration. In Fig. 4b, we present the reference free energies per particle βf_{ref} as function of the reduced density, computed using the two methods: the third order virial approximation agrees well with the TI, except, once again, at high density for the lowest salt concentration. As shown in Section 3.3, the critical points always lies well within the dilute regime,

where all the methods considered are bound to predict the same reference free energy. This strongly confirm the validity of the numerically-undemanding approach adopted in Ref. [13].

3.3. Critical points and coexistence

We now compute the critical points using as excess free energy either the virial approximation (Eq. (6)) or TI (Eq. (3)).

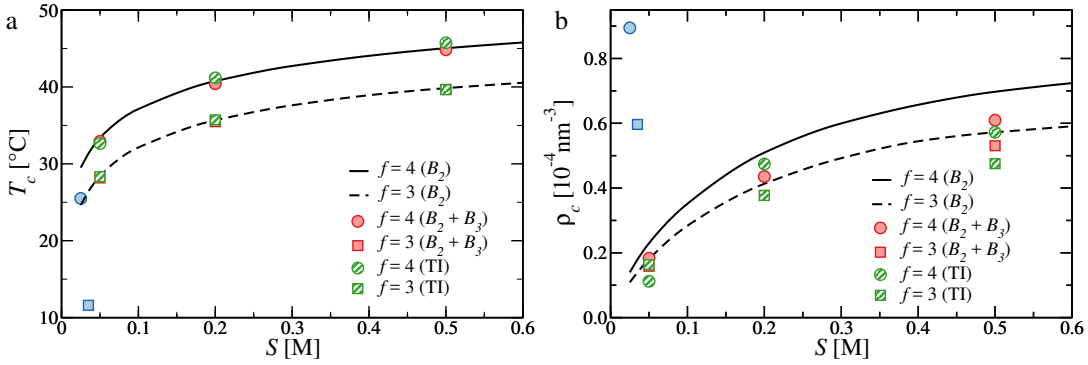

Figure 5. a) Critical temperature as function of salt concentration, b) Critical density as function of the salt concentrations. Circles and full line refer to tetramers ($f = 4$), squares and dashed line to trimers ($f = 3$). The data has been obtained with: the second order virial approximation [13] (black lines), the third order virial approximation (full red symbols), thermodynamic integration (dashed green symbols) and experiments [8] (full blue symbols).

In Fig. 5, we report critical temperatures (Fig. 5a) and critical densities (Fig. 5b) for tetramers and trimers. We report also the results of the second-order virial coefficient, as done in Ref. [13], for direct comparison. Note that even though we use the values of B_2 reported in Ref. [13], the two sets of results differ since the corrections to ΔH and ΔS due to the presence of dangling ends were miscalculated (see Appendix). Looking at the comparison, it is evident that the second-order virial expansion of the free energy is almost as accurate as the third-order and TI expressions in predicting the locus of the critical temperatures. Concerning the critical densities, the more accurate methods predict even smaller critical densities than the second order virial approximation; the disagreement is, however, small with respect to the discrepancy that still persists with respect to the experimental data. Such a poor agreement should be ascribed to a limitation of the WTPT, which is known to underestimate the critical density [24,32].

Next, we examine how employing more accurate expressions for the free energy affects the coexistence regions. In Fig. 6, we report the prediction of WTPT for tetramers at different salt concentrations (trimers exhibit the same qualitative trends). The main panel, which shows a comparison between the temperature-density phase diagram computed with the second (lines) and third order (symbols) virial approximations, reflects the qualitative trends noted before for the critical point. As expected, the range of temperatures remains pretty much unchanged, while the coexistence region shrinks towards smaller densities. From the inset of Fig. 6, we can appreciate that the low density branch remains untouched, since at low density the second-order virial expansion is already very accurate, while at higher densities the third term of the virial increases the excess free energy, thereby stabilising the liquid.

Lastly, we look at the effect of the tail correction of the free energy [14] to the critical temperature, shown in Figure 7. The increased effective repulsion does not significantly change the critical densities, which is therefore not shown. At low salt concentration the electrostatic repulsion shifts T_c by as much as 8 K for both tetramers and trimers, improving the agreement with experimental results. Upon increasing S the effect weakens and becomes effectively irrelevant above $S \approx 0.3$ M. The crossing of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 March 2018 doi:10.20944/preprints201803.0203.v1

Peer-reviewed version available at Polymers 2018, 10, 447; doi:10.3390/polym10040447

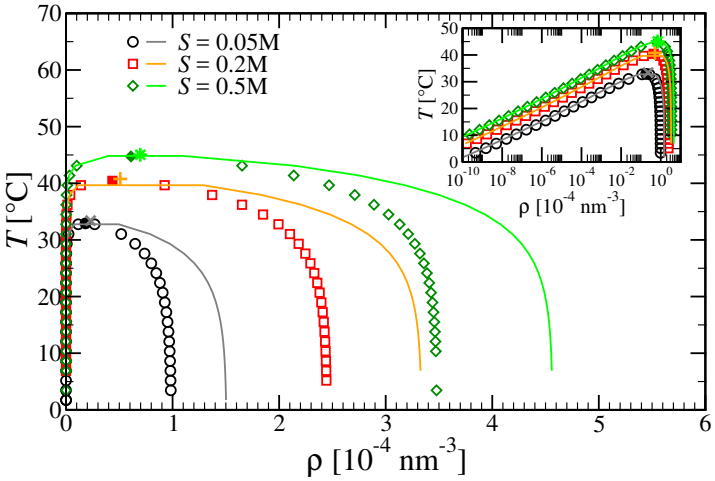

Figure 6. Main panel: Coexistence regions in the $T - \rho$ plane for tetramers at different salt concentrations. Symbols and lines refer to data obtained with: second order virial approximation (full lines), third order virial approximation (open symbols); full symbols refer to critical points. Inset: as in main panel, semi-log scale emphasizes the low density branch.

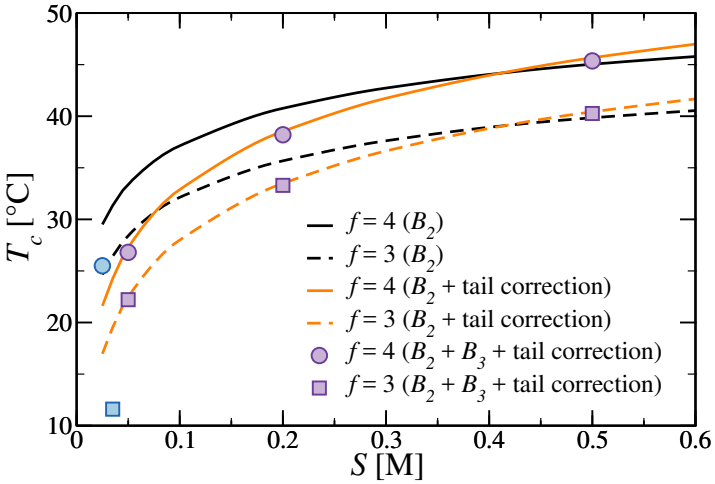

Figure 7. Critical temperature as function of salt concentrations, estimated by taking into account the contribution of the the repulsive tails, as done in Ref. [14]. Circles and solid lines refer to tetramers ($f = 4$), squares and dashed lines to trimers ($f = 3$). The data has been obtained with: the second order virial approximation [13] (orange and black lines for data obtained with and without tail corrections), the third order virial approximation (full violet symbols) and experiments [8] (full blue symbols). Data obtained with thermodynamic integration overlap with the $B_2 + B_3$ symbols and are therefore omitted for clarity.

lines obtained with and without corrections is most likely a spurious effect due to the salt-dependence of the entropy change we assume (see Appendix for details). Improving these estimates requires a better understanding of the tail effect on the DNA hybridisation free energy for different secondary structures and local geometries, calling for additional experiments and simulations.

4. Conclusions

We have reported a computational study aimed at accurately evaluating the phase behaviour of all-DNA systems. This was done by employing the hybrid theoretical/numerical approach introduced

in Ref. [13], where numerical simulations are used to calculate the inputs required by the Wertheim thermodynamic perturbation theory to evaluate the system free energy. We performed extensive numerical simulations to precisely compute one of these inputs, namely the reference free energy, for trivalent and tetravalent DNA nanostars at different temperatures, densities and salt concentrations. We have used this data to perform a close comparison with the results obtained through a much simpler virial expansion truncated at the second order, as done in Ref. [13].

The numerical results reported here show that such a second virial coefficient approximation faithfully reproduces the behaviour of the pressure in the reference fluid in a remarkably large interval of densities (essentially up to the semi-dilute regime). The agreement we observe demonstrates that, indeed, the effective pair-wise interaction is sufficient to reproduce the reference fluid structure and provides an inexpensive and accurate route to estimate the pressure, even at much higher density. Next, we have used the pressure data to estimate the reference free energy necessary to Wertheim TPT by truncating the virial expansion to the third term and by performing thermodynamic integration. The critical points computed with the WTPT exhibit a good agreement between each other and with the much simpler second-order virial approximation, once again demonstrating the validity of the approach developed in Ref. [13]. The only notable difference is a shift of the density of the coexisting liquid towards smaller values, which is to be ascribed to the increased repulsion with respect to the second-order virial expansion.

We have also looked at the effect of including an additional correction to the free energy due to the presence of nearby non-binding nucleotides to the hybridisation of the sticky ends, estimated by extrapolating values taken by Ref. [14]. This tail contribution effectively reduces the critical temperature, improving the agreement with experiments [8].

To conclude, we have shown that the method developed in Ref. [13] is not only very accurate despite its simplicity, but also easily extensible by incorporating additional terms into the WTPT theoretical description of the system. The agreement between the theoretical and experimental results shows that this approach already allows for a simple yet accurate *in silico* evaluation of the phase behaviour of all-DNA systems, providing a useful tool to assist future experimental realisations of all-DNA materials. [1]

Appendix Calculation of the hybridisation free energy

In general, the hybridisation free energy at any temperature is given by

$$\Delta G(T) = \Delta H + T\Delta S \tag{A1}$$

where ΔH and ΔS are the difference in enthalpy and entropy, that are assumed to be T-independent. From SantaLucia, the free energy at 37 degrees is given by:

$$\Delta G(37^\circ) = \Delta G_{\text{init}}(37^\circ) + \Delta G_{\text{sym}}(37^\circ) + \sum \Delta G_{\text{stack}}(37^\circ) + \Delta G_{\text{term}}(37^\circ) + \Delta G_{\text{dangl}}(37^\circ) \tag{A2}$$

The sequence we are interested in is $(5'A) - CGATCG - 3'$: this duplex is self-complementary, which means $\Delta G_{\text{symmetry}}(37^\circ) \neq 0$, and is attached to the arm of the nano-star through an unpaired A base at the $5'$ end.

$$\begin{array}{lccccccc} (5'A) - CGATCG - 3' = & \Delta G_{\text{sym}}(37^\circ) + & CG + & GA + & AT + & TC + & CG + & \Delta G_{\text{term}}(37^\circ) \\ 3 - GCTAGC - (A5') & & GC & CT & TA & AG & GC & \end{array} \tag{A3}$$

plus the additional contribution due to the dangling ends. SantaLucia gives the values for ΔH and $\Delta G(37^\circ)$ only. The entropy difference can be estimated by using Eq. (A1), as

$$\Delta S = \frac{\Delta H - \Delta G}{T} \tag{A4}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 March 2018 doi:10.20944/preprints201803.0203.v1

Peer-reviewed version available at Polymers 2018, 10, 447; doi:10.3390/polym10040447

from which, at $T = 37^\circ$ C and $S = 1$M, we can obtain the entropy difference ΔS.

Table A reports the SantaLucia contributions due to the stacking; the entropy difference has been estimated by using Eq. (A4). Notice that the $5'$ and $3'$ ends are reversed with respect to oxDNA convention. Data reported in Table A are considered accordingly.

Sequence	ΔH (kcal/mol)	ΔS (cal/K mol)	$\Delta G(37^\circ)$ (kcal/mol)
CG/GC	-9.8	-24.38	-2.24
GA/CT	-7.8	-21.02	-1.28
AT/TA	-7.2	-21.34	-0.58
TC/AG	-7.8	-21.02	-1.28
CG/GC	-9.8	-24.38	-2.24

The sum of all stacking contributions are:

$$\Delta H_{\text{stack}} = -42.4 \text{ kcal/mol} \qquad \Delta S_{\text{stack}} = -112.139 \text{ cal/ K mol}$$

The contribution for the initial G/C pair amounts to

$$\Delta H_{\text{init}} = 0.2 \text{ kcal/mol} \qquad \Delta S_{\text{init}} = -5.7 \text{ cal/ K mol}$$

The symmetry correction carries a contribution

$$\Delta H_{\text{sym}} = 0 \text{ kcal/mol} \qquad \Delta S_{\text{sym}} = -1.4 \text{ cal/ K mol}$$

Before the dangling end correction, the total enthalpy and entropy differences are $\Delta H = -42.2$ kcal/mol and $\Delta S = -119.23$ cal/ K mol. Finally, following SantaLucia, the sticky end sequence has two CA/G $3'$ dangling ends. Each dangling end gives a contribution of

$$\Delta H_{\text{dangl}} = -5.9 \text{ kcal/mol} \qquad \Delta S_{\text{dangl}} = -16.38 \text{ cal/ K mol}$$

The final free energy difference is therefore $\Delta H = -54$ kcal/mol and $\Delta S = -151.99$ cal/K mol.

We complement the calculation of the bonding free energy by also taking into account the effect of the nanostar on the bond formation, following Ref. [14]. As reported in the paper, the shift in the hybridization free energy $\delta\Delta G$ observed experimentally, due essentially to electrostatic repulsions, is independent of the duplex length, depends on the salt concentration S and reaches a constant value for sufficiently long tails ($n_{\text{tail}} \gtrsim 4$). Since the "tail" is, in our case, an entire double-stranded arm composed of 10 nucleotides, we consider the limiting (plateau) value of $\delta\Delta G$ at the proper salt concentration. We then consider Section S3.3 and Fig. S7 of the Supplementary Information of Ref. [14], in which a numerical estimate for the extra contribution to the hybridization enthalpy and entropy in presence of inert tails is provided. The data reported concern a duplex of length $n_{\text{hybrid}} = 9$ with tails of length $n_{\text{tail}} = 7$, and hence these numbers should be taken with a grain of salt. Rescaling the numerical $\delta\Delta G$ to the experimental values (at fixed salt concentration $S = 0.05M$) and rescaling the enthalpy and entropy shifts accordingly, we obtain that

$$\delta\Delta H_{\text{tail}} = 0.30296 \text{ kcal/mol} \qquad \delta\Delta S_{\text{tail}} = -0.00260 \text{ kcal/K mol} \tag{A5}$$

We assume that $\delta\Delta S_{\text{tail}} = \delta\Delta S_{\text{tail},0} + \alpha \ln(\text{S})$; such choice is motivated by the observed dependence of $\delta\Delta G_{\text{tail}}$ on the salt concentration, which we assume, following SantaLucia and for the lack of further experimental data, to have a purely entropic origin. Fitting the experimental data (Fig.3e of Ref. [14]) we obtain $\alpha = 1.6821$ cal/K and $\delta\Delta S_{\text{tail},0} = 2.4525$ cal/K. The shift in the Gibbs free energy for every temperature and every salt concentration is then given by

$$\delta\Delta G_{\text{tail}} = \delta\Delta H_{\text{tail}} + (\delta\Delta S_{\text{tail},0} + \alpha \ln(S))\, T \tag{A6}$$

Acknowledgments: The computational results presented have been achieved using the Vienna Scientific Cluster (VSC). The authors kindly acknowledge the Open Access Publishing Fund of the University of Vienna for fundings.

Author Contributions: EL performed the numerical simulations. All the authors designed the research and wrote the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

1. Seeman, N.C. DNA in a material world. *Nature* **2003**, *421*, 427.
2. Seeman, N.C. DNA Nanotechnology: Novel DNA Constructions. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 225–248.
3. Di Michele, L.; Eiser, E. Developments in Understanding and Controlling Self-Assembly of DNA-Functionalized Colloids. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3115–3129.
4. Jones, M.R.; Seeman, N.C.; Mirkin, C.A. Programmable Materials and the Nature of the DNA Bond. *Science* **2015**, *347*, 1260901.
5. Maye, M.M.; Kumara, M.T.; Nykypanchuk, D.; Sherman, W.B.; Gang, O. Switching binary states of nanoparticle superlattices and dimer clusters by DNA strands. *Nature Nanotech.* **2010**, *5*, 116–120.
6. Varrato, F.; Di Michele, L.; Belushkin, M.; Dorsaz, N.; Nathan, S.H.; Eiser, E.; Foffi, G. Arrested demixing opens route to bigels. *Proc. Natl. Acad. Sci.* **2012**.
7. Winfree, E.; Liu, F.; Wenzler, L.A.; Seeman, N.C. Design and Self-Assembly of Two-Dimensional DNA Crystals. *Nature* **1998**, *394*, 539–544.
8. Biffi, S.; Cerbino, R.; Bomboi, F.; Paraboschi, E.M.; Asselta, R.; Sciortino, F.; Bellini, T. Phase Behavior and Critical Activated Dynamics of Limited-Valence DNA Nanostars. *Proc. Nat. Acad. Sci.* **2013**, *110*, 15633–15637.
9. Bomboi, F.; Romano, F.; Leo, M.; Fernandez-Castanon, J.; Cerbino, R.; Bellini, T.; Bordi, F.; Filetici, P.; Sciortino, F. Re-entrant DNA Gels. *Nat. Commun.* **2016**, *7*, 13191.
10. SantaLucia Jr, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.
11. Wertheim, M.S. Fluids with Highly Directional Attractive Forces. I. Statistical Thermodynamics. *J. Stat. Phys.* **1984**, *35*, 19–34.
12. Wertheim, M.S. Fluids with Highly Directional Attractive Forces. II. Thermodynamic Perturbation Theory and Integral Equations. *J. Stat. Phys.* **1984**, *35*, 35–47.
13. Locatelli, E.; Handle, P.H.; Likos, C.N.; Sciortino, F.; Rovigatti, L. Condensation and demixing in solutions of DNA nanostars and their mixtures. *ACS nano* **2017**, *11*, 2094–2102.
14. Di Michele, L.; Mognetti, B.M.; Yanagishima, T.; Varilly, P.; Ruff, Z.; Frenkel, D.; Eiser, E. Effect of Inert Tails on the Thermodynamics of DNA Hybridization. *J. Am. Chem. Soc.* **2014**, *136*, 6538–6541. PMID: 24750023.
15. Bomboi, F.; Biffi, S.; Cerbino, R.; Bellini, T.; Bordi, F.; Sciortino, F. Equilibrium Gels of Trivalent DNA-Nanostars: Effect of the Ionic Strength on the Dynamics. *Eur. Phys. J. E* **2015**, *38*, 1–8.
16. Fernandez-Castanon, J.; Bomboi, F.; Rovigatti, L.; Zanatta, M.; Paciaroni, A.; Comez, L.; Porcar, L.; Jafta, C.J.; Fadda, G.C.; Bellini, T.; Sciortino, F. Small-Angle Neutron Scattering and Molecular Dynamics Structural Study of Gelling DNA Nanostars. *J. Chem. Phys.* **2016**, *145*.
17. Bianchi, E.; Capone, B.; Coluzza, I.; Rovigatti, L.; van Oostrum, P.D.J. Limiting the valence: advancements and new perspectives on patchy colloids, soft functionalized nanoparticles and biomolecules. *Phys. Chem. Chem. Phys.* **2017**, *19*, 19847–19868.
18. http://dna.physics.ox.ac.uk. (Accessed Feb 1, 2017).
19. Doye, J.P.; Ouldridge, T.E.; Louis, A.A.; Romano, F.; Šulc, P.; Matek, C.; Snodin, B.E.; Rovigatti, L.; Schreck, J.S.; Harrison, R.M.; Smith, W.P. Coarse-Graining DNA for Simulations of DNA Nanotechnology. *Phys. Chem. Chem. Phys.* **2013**, *12*, 6253–6260.
20. Snodin, B.E.K.; Randisi, F.; Mosayebi, M.; Šulc, P.; Schreck, J.S.; Romano, F.; Ouldridge, T.E.; Tsukanov, R.; Nir, E.; Louis, A.A.; Doye, J.P.K. Introducing Improved Structural Properties and Salt Dependence into a Coarse-Grained Model of DNA. *J. Chem. Phys.* **2015**, *142*, 234901.

21. Russo, J.; Tartaglia, P.; Sciortino, F. Reversible gels of patchy particles: role of the valence. *The Journal of chemical physics* **2009**, *131*, 014504.
22. Rovigatti, L.; Šulc, P.; Reguly, I.Z.; Romano, F. A Comparison between Parallelization Approaches in Molecular Dynamics Simulations on GPUs. *J. Comput. Chem.* **2015**, *36*, 1–8.
23. Ciccotti, G.; Ryckaert, J. Molecular dynamics simulation of rigid molecules. *Comp. Phys. Rep.* **1986**, *4*, 346 – 392.
24. Rovigatti, L.; de las Heras, D.; Tavares, J.; da Gama, M.T.; Sciortino, F. Computing the Phase Diagram of Binary Mixtures: a Patchy Particle Case Study. *J. Chem. Phys.* **2013**, *138*, 164904.
25. Bianchi, E.; Largo, J.; Tartaglia, P.; Zaccarelli, E.; Sciortino, F. Phase Diagram of Patchy Colloids: Towards Empty Liquids. *Phys. Rev. Lett.* **2006**, *97*, 168301.
26. SantaLucia, J. A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. *Proc. Nat. Acad. Sci.* **1998**, *95*, 1460.
27. Reinhardt, A.; Frenkel, D. DNA Brick Self-Assembly with an Off-Lattice Potential. *Soft Matter* **2016**, *15*, 20395–20414.
28. Watzlawek, M.; Löwen, H.; Likos, C.N. The anomalous structure factor of dense star polymer solutions. *Journal of Physics: Condensed Matter* **1998**, *10*, 8189.
29. Wang, L.; Bloomfield, V.A. Osmotic pressure of semidilute solutions of flexible, globular, and stiff-chain polyelectrolytes with added salt. *Macromolecules* **1990**, *23*, 194–199.
30. Likos, C.N. Effective interactions in Soft Condensed Matter Physics. *Physics Reports* **2001**, *348*, 267–439.
31. Locatelli, E.; Capone, B. *Design of Self-Assembling Materials, Chapter 1*; Springer International Publishing, 2018.
32. Bianchi, E.; Tartaglia, P.; Zaccarelli, E.; Sciortino, F. Theoretical and Numerical Study of the Phase Diagram of Patchy Colloids: Ordered and Disordered Patch Arrangements. *J. Chem. Phys.* **2008**, *128*, 144504.