



Cite this: *Phys. Chem. Chem. Phys.*,  
2016, **18**, 6373

## Theory and simulation of DNA-coated colloids: a guide for rational design

Stefano Angioletti-Uberti,<sup>†a</sup> Bortolo M. Mognetti<sup>†b</sup> and Daan Frenkel<sup>\*c</sup>

By exploiting the exquisite selectivity of DNA hybridization, DNA-coated colloids (DNACCs) can be made to self-assemble in a wide variety of structures. The beauty of this system stems largely from its exceptional versatility and from the fact that a proper choice of the grafted DNA sequences yields fine control over the colloidal interactions. Theory and simulations have an important role to play in the optimal design of self assembling DNACCs. At present, the powerful model-based design tools are not widely used, because the theoretical literature is fragmented and the connection between different theories is often not evident. In this Perspective, we aim to discuss the similarities and differences between the different models that have been described in the literature, their underlying assumptions, their strengths and their weaknesses. Using the tools described in the present Review, it should be possible to move towards a more rational design of novel self-assembling structures of DNACCs and, more generally, of systems where ligand–receptor are used to control interactions.

Received 14th November 2015,  
Accepted 11th January 2016

DOI: 10.1039/c5cp06981e

www.rsc.org/pccp

<sup>a</sup> International Research Centre for Soft Matter, Beijing University of Chemical Technology, 100029 Beijing, P. R. China

<sup>b</sup> Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, Code Postal 231, Campus Plaine, B-1050 Brussels, Belgium

<sup>c</sup> Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, UK. E-mail: df246@cam.ac.uk

<sup>†</sup> These authors contributed equally to this work.

### 1. Introduction

Colloidal suspensions are well described by the same statistical mechanical equations as systems of atoms or small molecules. As a consequence, colloids behave in many respects as ‘scaled-up’ models of atoms, and like atoms, colloidal suspensions can be found in phases that resemble gases, liquids or crystals. There are, however, important differences. One is that colloids move in a solvent, rather than in vacuum. The other is that, through modification of the colloid or the solvent, we can tune the



**Stefano Angioletti-Uberti**

*Stefano Angioletti-Uberti: after completing his Master in Materials Science in Milan (Italy) in 2007, Stefano Angioletti-Uberti received his PhD in the same subject at Imperial College London (United Kingdom) in 2010. Following a postdoc in Cambridge, where he worked on the theory and simulation of DNA-coated colloids in the group of Daan Frenkel, he then obtained an Alexander von Humboldt Postdoctoral Fellowship to work at the Humboldt*

*University Berlin and the Helmholtz Centre Berlin (Germany). In September 2015, he was appointed Professor of Soft Matter at the International Research Centre for Soft Matter at the Beijing University of Chemical Technology (China).*



**Bortolo M. Mognetti**

*Bortolo Matteo Mognetti: after completing his studies in Milan (Italy), Bortolo Matteo Mognetti held post-doctoral positions in Mainz (Germany) and Oxford (United Kingdom) working in the field of computational soft matter. Following a post-doctoral position in Cambridge in the group of Daan Frenkel to study DNA coated colloids, in October 2013 he was appointed lecturer (chargé de cours) at the Université Libre de Bruxelles (Belgium).*



interactions between colloids in a way that is not possible for atoms or small molecules. Due to our ability to control colloidal interactions, colloids exhibit a much richer phase behaviour than atomic systems. Over the past two decades, our ability to design colloidal interactions has undergone a quantum jump with the development of so-called DNA-coated colloids (DNACCs), colloidal particles functionalised with short sequences of single-stranded DNA (ssDNA). DNA allows the colloids to bind selectively, either directly or through single-stranded linkers, to other particles or surfaces coated with complementary ssDNA sequences.

DNA-coated colloids were developed by the groups of Mirkin and Alivisatos<sup>1,2</sup> in the 1990s. Over the past decades, DNACCs (sometimes referred to as “spherical nucleic acid”<sup>3</sup>) have been studied intensively because of their promise for the development of new classes of ordered colloidal materials (see *e.g.* ref. 4–7). Furthermore, this system has interesting biomedical applications as ultra-sensitive biomarkers, detectors or efficient drug- and gene-delivery carriers.<sup>8–10</sup> In all cases, applications exploit the selectivity of DNA–DNA hybridization between a complementary pair, as well as specific features of the physics of multivalent interactions.

Several reviews on DNACCs have appeared in the literature, both from an experimental<sup>11,12</sup> and from a theoretical<sup>13,14</sup> perspective. Very recently, Jones have provided a general overview describing the use of DNA-mediated interactions for the programmable self-assembly of nanomaterials,<sup>15</sup> highlighting the connections, and the differences, between DNACCs and DNA-based nanotechnology.

The focus of the present Perspective is different. It is not our aim to provide a comprehensive overview of all simulation and modelling results in the field, but rather, to present a critical assessment of the various theoretical and computational approaches that have been proposed to describe DNACCs.

However, this Perspective is not just a Review: to illustrate the points that we make, we also include new results. We aim to highlight the connections between different theoretical models, and explain the differences between various coarse-graining strategies used to describe DNACCs. We believe that the present perspective is timely, as the rational design of novel DNACC-based materials will require a complete integration of theory and experiment.

At present, theory and computational work are mostly used to explain experimental observations *a posteriori*. Yet these experiments are still largely driven by empirical or semi-empirical rules.<sup>16</sup> Even when model predictions preceded experiments, the link between theory and experiments is often less than clear.

At present, there is a bewildering variety of DNACC models in the literature, and this very fact makes it difficult for the non-expert to decide what approach to use. The net result is that most models are never used by experimentalists. More worrying is the fact that theoretical models that have been shown to be based on questionable premises continue to be used in the literature, thus carrying the risk that analysis of the experimental data is flawed. For these reasons, the aim of this review is to discuss the various models in terms of their underlying assumptions, their range of applicability and their relative strengths and weaknesses, so that experimentalists can use them in an informed way.

The remainder of the paper will be structured as follows. In Section II, we describe the general principles common to the various analytical theoretical models, and then proceed to discuss the connections between them. In particular, we aim to establish a ‘hierarchy of accuracy’ in terms of the effects included. These analytical models contain input parameters that depend on the molecular structure of the system, *e.g.* the specific DNA-sequence grafted on the colloids, or details of the spacer used for grafting. When a systematic study of the effect of these parameters is required, and in order to relax some simplifying assumptions necessarily included in the analytical models, one should resort to computer experiments, *i.e.* simulations. We will deal with these “experiments” in Section III, where we will focus on the various coarse-graining strategies that have been proposed in the literature. In particular, we discuss which microscopic features are retained in the various models, and which are ignored. In particular, we will focus on the question if the model can yield meaningful predictions of the relevant experiments. In the Concluding section (Section IV), we will identify possible new directions where the existing theoretical and computational framework could be profitably integrated with experiments to yield a more rational design of DNACC-based systems.

## II. Theory

The common idea behind basically all theoretical models to describe DNA-mediated interactions in DNACCs<sup>17–27</sup> is to account correctly for the interactions induced by the DNA strands that coat the colloids (a schematic of the system is depicted in Fig. 1).

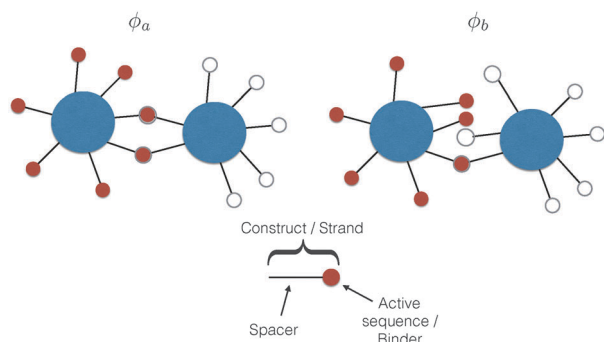


**Daan Frenkel**

*Daan Frenkel: received his PhD in Physical Chemistry from the University of Amsterdam (NL). He worked as a postdoc at UCLA and, subsequently, at Shell Research (Amsterdam), the Universities of Utrecht and Amsterdam and the FOM Institute AMOLF in Amsterdam. In 2007 he was appointed 1968 Professor of Theoretical Chemistry at Cambridge University. His research focuses on numerical simulations of many-body systems, with emphasis on*

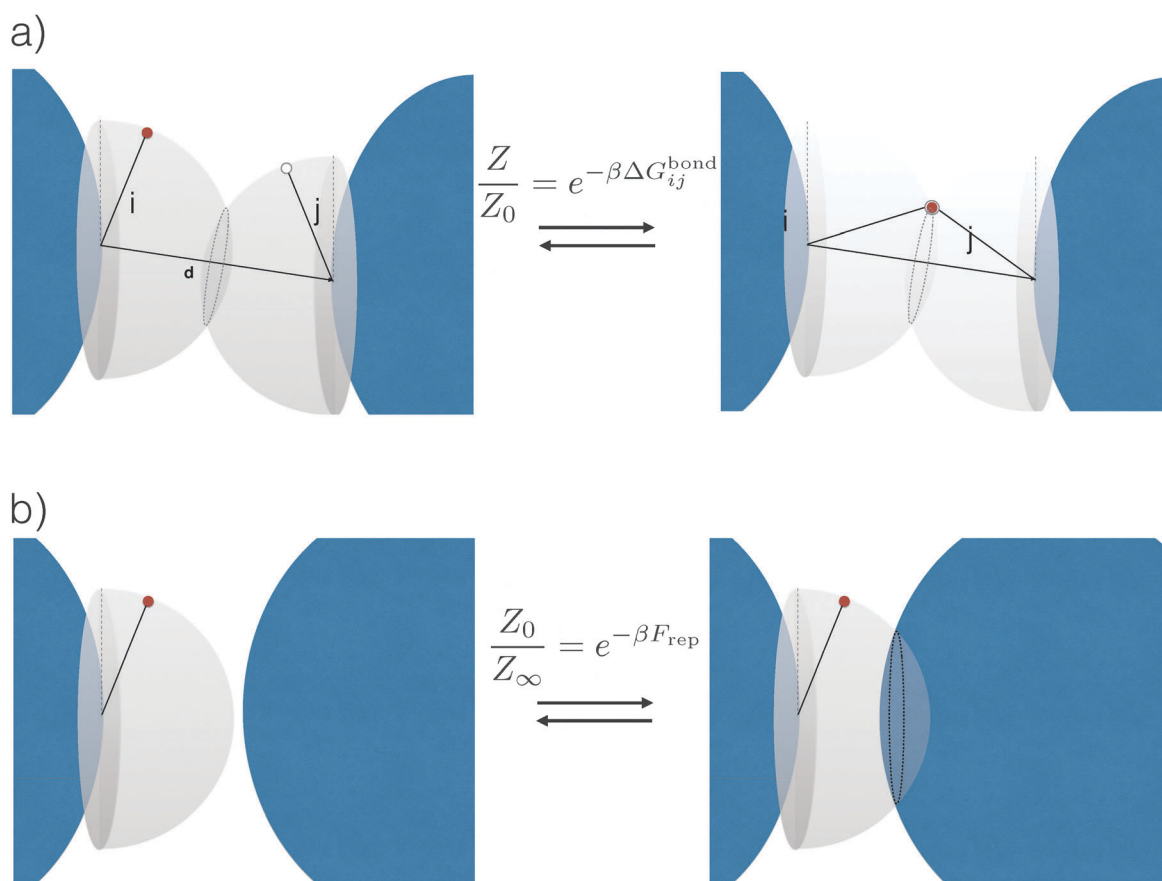
*problems relating to ordering and self-assembly. He is a member of six learned societies and has received numerous international awards and honorary degrees and professorships. He co-founded the International Research Centre for Soft Matter at the Beijing University of Chemical Technology (China).*





**Fig. 1** A schematic representation of a typical DNA-coated colloid, and the working principle behind their induced attraction. In DNACCs, single-stranded DNA, here represented as a red or white circle, is grafted onto a colloidal core via a spacer, typically another polymer or double-stranded DNA (dsDNA). Upon recognition of a complementary sequence, a single-stranded DNA can hybridise with it to form a joining dsDNA helix, leading to a highly-specific ligand–receptor bond. The figure presents two possible, mutually exclusive binding configurations,  $\phi_a$  and  $\phi_b$ . Considering all possible binding configurations and their energy provides the binding thermodynamics of the system.

On the one hand, as in general for all colloidal systems stabilised by (non-absorbing) grafted polymers, a non-specific repulsion arises from compression of the DNA strands trapped between the colloidal cores. On the other hand, a highly specific inter-particle attraction occurs because of the decrease in free-energy due to bond formation<sup>1,2</sup> (Fig. 2 provides a schematic illustration). That bond-mediated interactions are the driving force leading to aggregation is also evident from experimental studies that show no aggregation between DNACCs grafted with non-complementary strands.<sup>1,2,25</sup> In fact, other (non-specific) attractive interactions, in particular those arising from the van der Waals forces between the colloidal cores, are typically much weaker than those arising from bond formation on the length-scales at which bond-mediated binding occurs. This is mainly due to the fact that the non-specific steric repulsive forces due to DNA compression prevent the colloids to approach at short distances where dispersion forces are strong.<sup>26</sup> For similar reasons, the length-scale for binding, *i.e.* the typical distance between colloidal cores between bound particles, is dictated by the length of the spacer used to graft DNA on the surface of



**Fig. 2** Schematic representation of the origin of attractive and repulsive forces, here shown for a simple system of strands considered as rigid rods and in the case where only a single bond can be formed. (a) Upon binding, the system acquires an additional energy  $\Delta G_{ij}^{\text{bond}}$ , which depends on both the chemical identity of the active strands (white and red circles) as well as on details of the spacer (black sticks) and position of the tethering points (see eqn (3)). In particular, note that whereas in the unbound state (left) the active strands can freely move on a hemisphere, upon binding (right) they are forced to stay on a circle: this reduction in configurational space is at the origin of  $\Delta G^{\text{cnf}}$ , (see eqn (4)). (b) When all colloids are far apart from each other, the active strands can span a whole hemisphere. However, this space is again reduced once colloids start to get closer. This reduction is at the origin of the dominant repulsive forces  $F_{\text{rep}}$  between DNACCs.



the colloids, which is typically in the range of nanometers to tens of nanometers.<sup>17</sup> In various practical realisations of DNACCs, other polymer chains than DNA are grafted to the colloidal surface. Using an additional different polymer for steric stabilisation makes it possible to tune repulsion and attraction separately. Both theory and experiments have shown that tuning the relative strength of attraction and repulsion can lead to self-assembly of aggregates of different types.<sup>28–30</sup>

It is crucial to note that the interaction between two grafted colloid is due to both energy and entropy, the latter accounting for the fact that often not just one but several binding configurations are possible (see Fig. 1). Hence, to compute the interaction strength, we need to evaluate a free energy, rather than an energy, as would be the case for atoms in vacuum. To compute the free energy of interaction of DNACCs, we start from the statistical mechanical relation between the free energy  $F$  of a system and its partition function  $Z$ . We consider an arbitrary number of colloids, each of which has an arbitrary number of strands grafted on its surface. We choose a reference state where all colloids are at infinite distance from each other, and calculate the free-energy with respect to this state:

$$\begin{aligned}\beta F_{\text{interaction}} &= -k_{\text{B}}T \log \frac{Z(\{\mathbf{R}\})}{Z_{\infty}} \\ &= -k_{\text{B}}T \log \frac{Z(\{\mathbf{R}\})}{Z_0(\{\mathbf{R}\})} - k_{\text{B}}T \log \frac{Z_0(\{\mathbf{R}\})}{Z_{\infty}} \quad (1) \\ &= F_{\text{att}} + F_{\text{rep}}\end{aligned}$$

where  $Z = Z_0 + Z_{\text{bound}}$  is the partition function of all strands in the system (*i.e.* a sum over Boltzmann weights), which depends on their grafting points  $\{\mathbf{r}\}$ , as well as on the positions of all colloids  $\{\mathbf{R}\}$ .  $Z_0$  is the partial sum over all spacer states where no bond forms,  $Z_{\text{bound}}$  the partial sum over all states with at least one bond, and finally  $Z^{\infty} \equiv Z(\{\mathbf{R} = \infty\})$  is the partition function when all colloids are at infinite separation. With this splitting, we can identify two terms,  $F_{\text{att}}$  and  $F_{\text{rep}}$ , *i.e.* the attractive and repulsive component of the interaction, respectively. Note that when binding between strands on the same particle can occur, this identification cannot be made, although the general approach to calculate  $F_{\text{interaction}}$  (see eqn (1)) still holds. We stress that no approximation has been made in the derivation of eqn (1).

Since  $Z = Z_0 + Z_{\text{bound}}$ ,  $F_{\text{att}}$  is always negative (and, hence, attractive). Instead  $F_{\text{rep}}$  is the ratio between two partition functions where bonds are never allowed, although in one case ( $Z_0$ ) strands can feel the excluded volume due to all colloids, including the one on which they are grafted, whereas in the other ( $Z^{\infty}$ ) strands can only feel the latter. If we assume that only repulsive interactions between the colloidal cores and the strands occur, *i.e.* we consider non-adsorbing polymers, a good approximation for most type of polymers used in DNACCs, then  $Z_0/Z^{\infty} < 1$ , and  $F_{\text{rep}}$  is always positive (*i.e.* repulsive).

Use of eqn (1) would allow one to calculate the interaction between DNACCs including all type of many-body interactions arising from bond formations, and for arbitrary position of the colloids. This expression does not assume that DNA-mediated

interactions are pairwise additive – and, in fact, often they are not.<sup>28,31,32</sup> For this reason, the general expression is needed to compute the interaction free energy of dense phases of DNACCs.

Let us next consider how  $F_{\text{att}}$  and  $F_{\text{rep}}$  can be calculated. One route is both obvious and pointless: ‘brute-force’ atomistic simulations. Simulations that would describe the colloids, the DNA and the solvent in atomistic detail are simply much too expensive to be used to predict, say, phase diagrams. Fully atomistic simulations can certainly be used to study certain aspects of DNACCs, but not to study a suspension containing hundreds or thousands of such particles. In what follows, some degree of coarse-graining is therefore essential. Typically, this means that we will use an implicit solvent model and that we will assume that the binding free energy between complementary DNA strands is known. In other words: models to describe systems consisting of many DNACCs are always coarse-grained.

To arrive at an analytically tractable coarse-grained model, the following assumptions are commonly made: (1) we usually ignore the spatial extent of the nucleotide sequences that can hybridise to bind two strands together. The hybridising strand is then represented as a point-like attractive site, also called ‘active site’ or ‘sticky end’, tethered to the colloidal surface by a long chain – the spacer or linker<sup>17,19–27</sup> (see also Fig. 1). Representing the ‘sticky’ binding sequence as a point is a reasonable approximation as long as the spacer is much longer than the sticky end, a situation typically encountered in experimental systems.<sup>1,2,5,17,21</sup> Considering extended sticky ends results in slightly different binding affinities between constructs.<sup>23,33</sup> Accounting for the finite spatial extent for sticky end does not qualitatively change the theoretical treatment of DNACCs, and hence we limit our discussion to point-like sticky ends. (2) We ignore steric interactions between different strands and, more generally, any non-selective strand–strand interaction. This approximation is very common, but it is likely to break down for dense DNA ‘brushes’.

With the above assumptions,  $F_{\text{rep}}$  can be expressed as:

$$F_{\text{rep}} = -k_{\text{B}}T \sum_i \log \frac{\Omega_i}{\Omega_i^{\infty}}, \quad (2)$$

where the index  $i$  runs over all strands in the system, and  $\Omega_i$  and  $\Omega_i^{\infty}$  are, respectively, the partition function of a single linker when the colloids are in positions  $\{\mathbf{R}\}$  and the partition function for the case where all colloids are at infinite separation. For specific geometries and simple polymer models for the linker, such as rigid rods or gaussian chains, there are either exact or very accurate approximate expressions to calculate  $F_{\text{rep}}$ .<sup>21,23,26</sup> For more complex cases,  $F_{\text{rep}}$  can be obtained using standard Monte Carlo techniques to sample polymer conformations.<sup>24,25</sup> Rogers and Crocker showed that a simple algorithm that assumes that polymers can be described as ideal, non-interacting chains that cannot intersect the colloids gave essentially quantitative agreement with the repulsive interaction determined in experiments.<sup>25</sup> This agreement suggests that in this specific system, the neglect of chain–chain and attractive chain–colloid interactions is justified. It also suggests that electrostatic interactions due to the dense clouds of counter ions around DNA can be included





in an effective way *via* their influence on the single-chain conformations. These depend on the chosen length of the segments in the freely-jointed chain, which is the quantity directly affected by electrostatics.<sup>34</sup> In Section III we discuss non-specific interactions between spacers in more detail.

The calculation of  $F_{\text{att}}$  is more challenging, since here we need to account for all possible binding configurations with an arbitrary number of bonds (see Fig. 1). In order to do that, it is useful to split the problem of calculating  $F_{\text{att}}$  into two parts. The first stage involves calculating the bond formation energy for all individual pairs of  $(i, j)$  strands, and the second stage accounts for the fact that there may be many ways in which the DNA strands on adjacent colloids may be connected (as illustrated in Fig. 1). If the spacer does not influence the structure of the sticky end, one can split the bond formation free energy into two terms:<sup>21,23,26,28</sup>

$$\Delta G_{ij}^{\text{bond}} = \Delta G_{ij}^0 + \Delta G^{\text{cnf}}(\mathbf{r}_i, \mathbf{r}_j, \{\mathbf{R}_i\}). \quad (3)$$

The first term on the r.h.s. of eqn (3) is the free-energy of binding for untethered, complementary DNA strands  $i$  and  $j$  in solution:  $\exp[-\beta\Delta G_{ij}^0] = K_{ij}^{\text{eq}}\rho^0$ , where  $K_{ij}^{\text{eq}}$  denotes the equilibrium binding constant, and  $\rho^0$  the standard concentration of 1 M. The second term represents the configurational entropic cost associated with linking two tethered strands  $i$  and  $j$ : the number of configurations of a linked  $ij$  strand is less than that of the unlinked  $i$  and  $j$  strands<sup>35</sup> (see Fig. 2a).  $\Delta G^{\text{cnf}}(\mathbf{r}_i, \mathbf{r}_j, \{\mathbf{R}_i\})$  depends on the exact positions of the grafting points, and on the position of all nearby colloids in the system, since these exclude all strand configurations that would intersect them. Ref. 22 and 23 give an explicit expression that can be used to calculate this quantity for arbitrary linkers and colloids positions, which in its most general case reads:

$$\Delta G^{\text{cnf}}(\mathbf{r}_i, \mathbf{r}_j, \{\mathbf{R}_i\}) = -k_B T \log \frac{\Omega_{ij}(\mathbf{r}_i, \mathbf{r}_j, \{\mathbf{R}_i\})}{\Omega_{i,j}(\mathbf{r}_i, \mathbf{r}_j, \{\mathbf{R}_i\})\rho^0}, \quad (4)$$

where  $\Omega_{ij}$  is the partition function for two bound linkers  $i$  and  $j$ , whereas  $\Omega_{i,j}$  is the partition function for the unbound case. Both terms depend on  $\mathbf{r}_i, \mathbf{r}_j$  and  $\{\mathbf{R}_i\}$ , and can be calculated using coarse-grained models for the linkers.<sup>23</sup> When strand-strand interactions can be neglected, an equivalent and possibly more illuminating form of eqn (4) useful for calculations is:

$$\Delta G^{\text{cnf}}(\mathbf{r}_i, \mathbf{r}_j, \{\mathbf{R}_i\}) = -k_B T \log \left( p_{\text{ec}}(\mathbf{d}) \frac{W_{ij}}{W_i W_j \rho^0} \right) \quad (5)$$

where  $p_{\text{ec}}(\mathbf{d})$  denotes the fraction of all random configurations of the two chains tethered at two grafting points at separation  $\mathbf{d}$ , that have overlapping sticky ends, *i.e.* it is the probability that two non-interacting tethers will form a ‘bridging’ configuration. The terms  $W_{ij}$  and  $W_{i(j)}$  account for the effect of hard-core overlaps with the colloids: they denote the number of  $i, j$  or  $ij$  configurations that do not overlap with the colloidal cores (see Fig. 2). Note that,  $W_{i(j)}$  for the unbound linkers is equal to the  $\Omega_{i(j)}$ s appearing in the definition of  $F_{\text{rep}}$ , eqn (2).

$\Delta G^{\text{cnf}}(\mathbf{r}_i, \mathbf{r}_j, \{\mathbf{R}_i\})$  can be either calculated analytically for simple polymer models, or computed *via* Monte Carlo simulations, as we discuss in Section III. In a small number of (important) cases,

it can be approximated analytically.<sup>17,21,23</sup> The different theories that have been proposed to estimate  $F_{\text{att}}$  make different approximations to calculate this quantity. However, all models<sup>20–22,25</sup> assume that strands are non-interacting (although the interaction can still be included in calculating the bond energy<sup>23</sup>). This approximation is reasonable when different chains are not close to each other. At higher grafting densities (see *e.g.* ref. 31 and 32), the simple analytical approaches break down, but the simple theories can still be used to gain insight in the effect of the structure of the spacer (*e.g.* rigid rods or freely jointed chains) on the interactions between DNACCs. For example, “theory accounts” quite well for the experimentally observed trends in the melting properties of DNACCs.<sup>21,22</sup>

Under the assumption that the binding free energy  $\Delta G_{ij}^{\text{bond}}$  between  $i$  and  $j$  does not depend on the presence of other strands (*i.e.* the non interaction assumption), knowledge of  $\Delta G_{ij}^{\text{bond}}$  for all possible pairs of constructs  $(i, j)$  is sufficient to write down an exact expression for the partition function  $Z$ :

$$\frac{Z}{Z_0} = \sum_{\phi} \exp \left[ - \sum_{(i,j) \in \phi} \beta \Delta G_{ij}^{\text{bond}} \right] \quad (6)$$

where the sum runs over all binding configuration  $\phi$  of the system (see Fig. 1), *i.e.*  $Z/Z_0$  is simply equal to the sum of all Boltzmann weights due to bonds energies over all possible binding configurations. Note that  $\Delta G_{ij}^{\text{bond}}$  depends on the position of the strands *via*  $\Delta G^{\text{cnf}}$ , see eqn (3), and hence depends on the spatial configuration of the colloids.

An exact enumeration of the terms in eqn (6) becomes rapidly intractable since the number of binding configurations increases very rapidly with the number of strands in the system. Hence eqn (6) is typically simplified under some assumptions. It is at this level of approximation that differences between the various analytical theories show up.<sup>17,19,21,23,25,27</sup> Short of exact enumeration,  $Z/Z_0$ , and hence  $F_{\text{att}}$ , can be calculated to any given accuracy *via* Monte Carlo simulations using thermodynamic integration,<sup>36</sup> providing a route to fully include all possible competitive effects between formation of different binding configurations, as well as about the spatial distribution of the strands. The exact solution reads:<sup>23,24,26</sup>

$$\beta F_{\text{att}}(\beta \Delta G^0) = \int_{-\infty}^{\beta \Delta G^0(T)} \sum_{i < j} p_{ij}(\beta \Delta G^{0'}) d\beta \Delta G^{0'}, \quad (7)$$

where  $p_{ij}$  is the probability that a bond between strands  $i$  and  $j$  is observed, as sampled by MC simulations. In ref. 23 and 27 we have shown that the ‘exact’ Monte Carlo results are approximated extremely well by a very simple expression:

$$F_{\text{att}} = \sum_i \log p_i + \frac{1}{2}(1 - p_i), \quad (8)$$

and

$$p_i + \sum_j p_i p_j \exp(-\beta \Delta G_{ij}^{\text{bond}}) = 1, \quad \forall i, j \quad (9)$$

where the indexes  $i$  and  $j$  run over all possible strands in the system. In eqn (8),  $p_i$  is the probability that strand  $i$  is not bound,



found by solving the set of coupled self-consistent equations represented by eqn (9). Eqn (8) and (9) can be derived using a recursion formula to count all possible binding configurations (see ref. 27).

In our derivation of eqn (8) and (9) we used an approximate yet very accurate formula for counting the degeneracy of each binding configuration, accounting for the fact that no two linkers can bind to the same target.<sup>27</sup> From a combinatorial point of view, it turns out that our procedure can be mapped onto Wertheim's theory for the equilibrium properties of fluids with highly directional intermolecular forces.<sup>37–39</sup> Wertheim's solution is a basic ingredient of the widely used SAFT theory.<sup>40</sup> In practice, each reactive DNA strand (or, more generally, each binder) must be considered in the Wertheim picture as a spatially and orientationally fixed particle with a single binding site, but each having a unique bond energy, which in our case must also be defined as in eqn (3). In our system, we also have the additional benefit that some of the approximations Wertheim used are always exactly satisfied conditions, in particular the fact that if two strands are bound a third one cannot bind at the same time. Once the problem is cast in these terms, it becomes mathematically equivalent to that of Wertheim.

The approximations underlying eqn (8) and (9) become exact in the case of mobile constructs, see ref. 28 and Section III for more details. Finally, we stress that although these equations provide the  $F_{\text{att}}$  induced by the interaction between grafted strands. It is easy to show that they remain the same when binding between these strands is not direct, but is mediated by (or compete with), free complementary strands in solution, another widely used bonding scheme in DNACCs. In practice, eqn (8) and (9) remain valid, as well as the configurational part of the bond free-energy  $\Delta G^{\text{cnf}}$ , whereas the solution hybridisation free-energy  $\Delta G_{ij}^0$  between strands must be shifted with a density dependent factor.

Since eqn (8) and (9) agree extremely well with the corresponding MC simulations, they also properly reproduce all the (important) correlations introduced by the competition between strands for the same binding partner. Moreover, they are completely general in that they can be used to treat an arbitrary number of different binding pairs, with any spatial distribution. Hence, they can be used to calculate colloid–colloid interactions also when highly non-homogeneous DNA coatings are present, as in the experimental system reported by Pine and coworkers mimicking patchy particles.<sup>41</sup>

Earlier expressions that have been proposed in the literature<sup>17,19,21,22</sup> can be derived as approximations of eqn (8) and (9) under different assumptions.<sup>27</sup> In particular, the aforementioned models all shared a mean-field approximation for binding, *i.e.* unlike eqn (8) and (9) they do neglect correlations that are due to the fact that two strands cannot bind simultaneously to the same binding partner (the ‘valence constraint’<sup>23</sup>). Furthermore, strands were not only treated as if they were completely independent from each other, but also as if they were all statistically equivalent. The first approximation becomes progressively worse as the binding strength of bonds increases (relative to  $k_{\text{B}}T$ ). This drawback can be heuristically corrected by

a self-consistent mean field treatment,<sup>42</sup> which yields the mean-field version of eqn (9). The assumption that all bonds are statistically equivalent gets worse as the grafting of chains becomes more heterogeneous, since in this case each strand will experience a different environment. However, if we assume that both approximations are justified, then the expression for  $F_{\text{att}}$  simplifies:<sup>17,19,21,22</sup>

$$F_{\text{att}} \approx -k_{\text{B}}T\langle n \rangle \quad (10)$$

where  $\langle n \rangle$  is the number of bonds formed in the system. This expression for  $F_{\text{att}}$  is called the “weak binding” or the “Poisson” approximation.<sup>17,19,21,22,25</sup> It provides a lower bound for the real interaction energy, as it always overestimates the number of possible bonds: any correlation would reduce the number of bonds that can form. It has been pointed out<sup>25</sup> that models that are based on the Poisson approximation may overestimate the attractive interaction by as much as two orders of magnitude, corresponding to approximately  $\approx 5k_{\text{B}}T$  per DNA bridge. The “weak binding” approximation can be obtained formally from the expression given by eqn (7) by a simple assumption. The integrand in eqn (7) is nothing but the expected number of bound pairs in the system. If this number can be approximated by a Poisson distribution (an approximation that breaks down at higher binding strength<sup>24</sup>), then eqn (10) follows from eqn (7). It can be also shown that eqn (10) corresponds to a first-order approximation of eqn (8) and (9) (the second-order correction is positive), and its validity is limited to a region where each strand has very few binding partners or small binding energy (hence the “weak binding regime”).<sup>27</sup> Another possible drawback of eqn (10) is that it cannot be generalised easily to treat an arbitrary number of binding partners with different binding energies, since eqn (10) was derived considering not just independent, but also equivalent strands. This is a serious limitation, as the phase diagram of DNACCs can be ‘designed’ by making colloids with more than one type of strand (see *e.g.* ref. 29).

A possible way to estimate  $F_{\text{att}}$  for an arbitrary number of types of strands was proposed by Rogers and Crocker.<sup>25</sup> Heuristically keeping eqn (10) as the expression for the attractive contribution, they proposed to use mass balance equations valid for chemical reactions in solution to describe binding between grafted strands, an approach dubbed “Local Chemical Equilibrium” (LCE). In ref. 25 it is assumed that the binding for grafted strands can be treated as an equilibrium binding reaction in solution with a non-homogeneous distribution of active sites. This distribution is obtained by Monte Carlo sampling the position of the end-point of polymer chains.  $F_{\text{att}}$  is then approximated as:

$$C_{\alpha\beta}(\mathbf{r}) = \frac{C_{\alpha}(\mathbf{r})C_{\beta}(\mathbf{r})}{\rho^0} \exp(-\beta\Delta G_{\alpha\beta}^0) \quad (11)$$

$$C_{\alpha}(\mathbf{r}) \approx C_{\alpha}^0(\mathbf{r}) - \sum_{\beta} C_{\alpha\beta}(\mathbf{r}) \quad \forall \alpha \quad (12)$$

$$\beta F_{\text{att}} = - \sum_{\alpha < \beta} \int_V d\mathbf{r} C_{\alpha\beta}(\mathbf{r}), \quad (13)$$

where  $C_{\alpha(\beta)}(\mathbf{r})$  are the equilibrium local concentrations, of free, unbound active sites of type  $\alpha(\beta)$  (note that in this formalism all



strands with the same active site are considered equivalent, whereas in eqn (8) and (9) all strands have different identities)  $C_{\alpha(\beta)}^0(\mathbf{r})$ , the initial concentration of active sites when no binding occurs, and  $C_{\alpha\beta}(\mathbf{r})$  the concentration of bound  $\alpha$ - $\beta$  pairs once binding is allowed. Note that eqn (13) represents a very strong constraint: in general, mass conservation makes it valid when concentrations are integrated over the whole volume. Eqn (12) and (13) instead assume that this is true at each point in space, a much stronger statement. In fact, the approximation in eqn (13) is in general not justified,<sup>23</sup> as we discuss in more detail below.

### A. Equilibrium in bulk binary reactions

In order to derive the (LCE) expression for DNA binding, it is important to realise that tethered DNA strands are distinguishable. The derivation of the condition for local chemical equilibrium is therefore different from the case where the reactants and products comprise molecules that are indistinguishable. As an illustration, we will assume that the tethering of DNA strands results in a confinement of the 'reactive' ends, but that the free-energy cost of tethering can be ignored (this assumption can subsequently be refined). We consider the hybridisation reaction  $\alpha + \beta \rightleftharpoons \alpha\beta$ . As we can ignore the tethering free energy, we can treat the mixture of  $\alpha$ ,  $\beta$  and  $\alpha\beta$  strands as an ideal gas of distinguishable particles. We assume that,  $\alpha$ ,  $\beta$  and  $\alpha\beta$  are confined to a volume  $V$  that is small, but large enough to contain many molecules. The partition function of a system with  $N_\alpha$  ( $N_\beta$ ) monomers of type  $\alpha$  ( $\beta$ ) and  $N_{\alpha\beta}$  complexes is then:

$$Q(N_\alpha, N_\beta, N_{\alpha\beta}, V, T) = \frac{N_\alpha^0!}{N_\alpha!N_{\alpha\beta}!} \frac{N_\beta^0!}{N_\beta!N_{\alpha\beta}!} N_{\alpha\beta}! V^{N^0 - N_{\alpha\beta}} \times \prod_{x=\alpha, \beta, \alpha\beta} \left( \frac{q_{\text{int}}(x)}{A_x^{3N}} \right)^{N_x}, \quad (14)$$

where  $N_\alpha^0$  ( $N_\beta^0$ ) denotes the original (*i.e.* pre-reaction) concentration of  $\alpha$  ( $\beta$ ), and  $N^0 \equiv N_\alpha^0 + N_\beta^0$ , and  $q_{\text{int}}(x)$  is the internal partition function of complex  $x$  ( $x = \alpha, \beta, \alpha\beta$ ). Note that the combinatorial factors in eqn (14) are not related to indistinguishability but to the number of ways to select  $N_{\alpha\beta}$  reaction partners from  $N_\alpha$  and from  $N_\beta$  monomers. The factor  $N_{\alpha\beta}!$  accounts for the number of ways in which  $N_{\alpha\beta}$  distinguishable particles of type  $\alpha$  and the same number of type  $\beta$  can be paired. The partition function is a function of  $N_{\alpha\beta}$ . It is maximised for

$$\frac{N_\alpha N_\beta}{N_{\alpha\beta} V} = \frac{C_\alpha C_\beta}{C_{\alpha\beta}} = \exp\left(-\beta \Delta G_{\alpha\beta}^0\right) / \rho^0. \quad (15)$$

Hence, for large enough numbers of reactive particles in the volume  $V$ , the hybridisation of tethered strands obeys the same equilibrium expression as a binary gas mixture. However, for small particle numbers this expression fails even qualitatively (see Appendix A). This means that the LCE is reasonable in the 'mean-field' limit where any given DNA strand can bind to many others, but it fails badly when this number is  $\mathcal{O}(1)$  or less. This is typically the case when the average distance between grating points is not small compared with the length of the tether.

It may seem surprising that the approach of ref. 25 yields results that were compatible with the experimental data, given the

fact that eqn (12) is based on an approximation, and that the mean-field assumption is not expected to be particularly good in the case studied in ref. 25. It is probably more likely that the accuracy of the experiments reported in ref. 25 was insufficient to discriminate between different models.<sup>23,43</sup>

For this reason, the best way to discriminate between various approximate models is to use computer 'experiments' in which 'exact' Monte Carlo simulations of a specific model are used to compare between different approximations. In particular, if a model cannot reproduce the behaviour of a well-characterised model system, its agreement with experiments must be fortuitous. In ref. 24, we compared the different analytical theories with MC calculations to test the validity of mass balance equations and found that LCE can give results that differ substantially from the 'exact' simulation results, especially at high binding strengths (see also Fig. 3). It should come as no surprise that the agreement worsens at high binding energies as eqn (10) is only recovered from our analytical theory eqn (8) in the weak-binding limit  $p_i \rightarrow 1$ .<sup>27</sup>

A more positive finding is that, in the weak-binding limit eqn (10) remains still valid for an otherwise arbitrary distribution of binding strengths. No such good news is in store for the LCE: it fails, even in the weak-binding limit because mass-balance equations can only be used to calculate the bond densities when considering the binding between free-strands in solution, not for grafted strands, where the local mass-conservation implicit in eqn (12) does not hold. However, starting from our self-consistent equations, eqn (8) and (9) we can show that a better LCE approximation for the binding densities is given by:

$$C_\alpha^0(\mathbf{r}) = C_\alpha(\mathbf{r})/p_\alpha \quad (16)$$

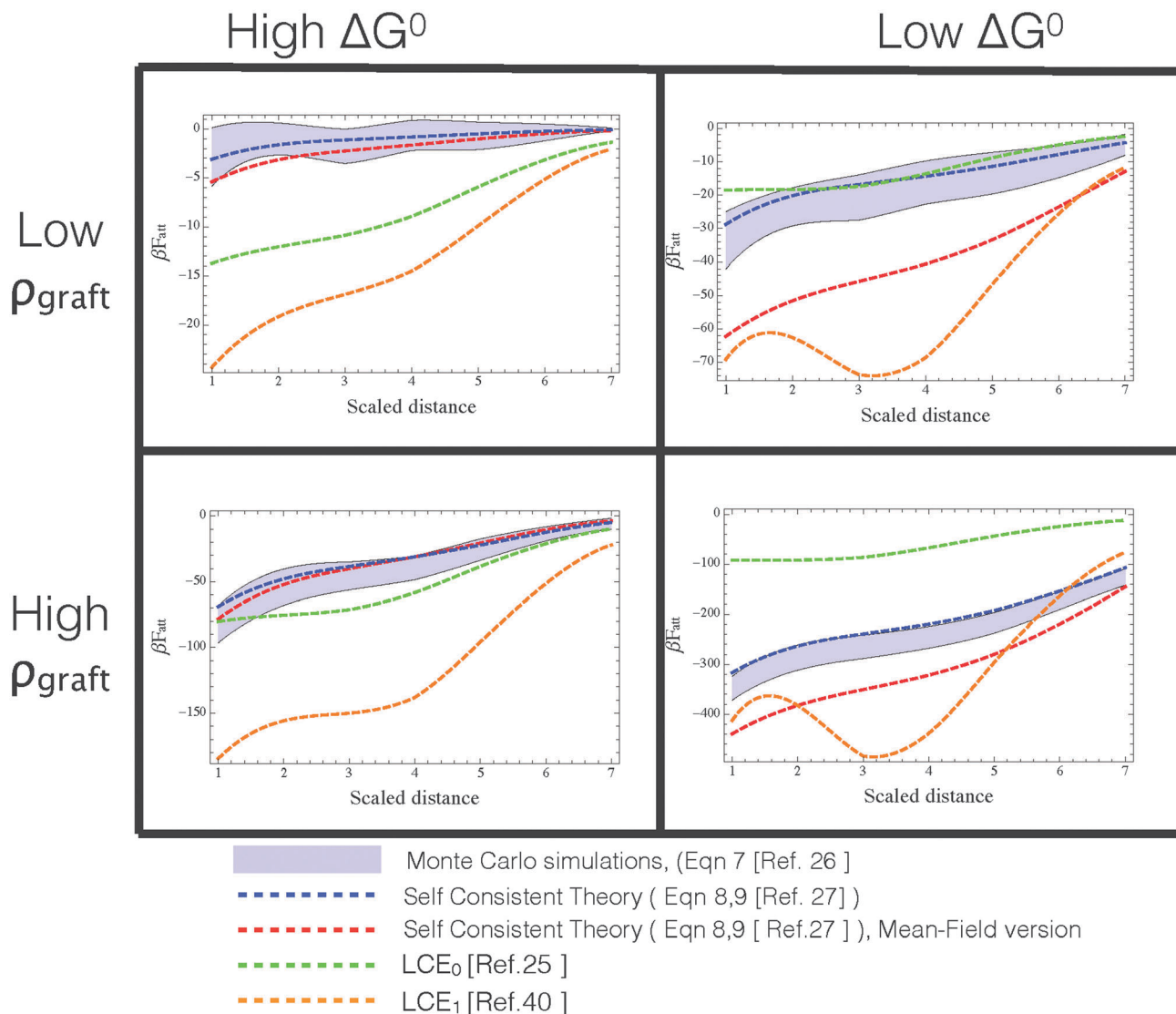
together with the self-consistent condition:

$$\int d\mathbf{r} C_\alpha(\mathbf{r}) + \sum_\beta C_{\alpha\beta}(\mathbf{r}) = N_\alpha, \quad (17)$$

$N_\alpha$  being the number of strands of type  $\alpha$  in the system. Once corrected, the LCE treatment gives essentially exact results, at least for weak binding where eqn (10) is valid, when compared with full Monte Carlo simulations, unless the number of binding partners per linker is very small (see Appendix A and Fig. 3). The advantage of eqn (8) and (9) is that these equations agree with the MC simulations over a wide range of binding regimes, not just in the weak binding limit (see Fig. 3). In this figure, we also show a comparison with a form of LCE that was recently proposed by Rogers and Manoharan. The latter theory assumes local mass balance but uses a different approximation for  $F_{\text{att}}$  (ref. 44, ESI).

The theoretical work presented thus far is based on various assumptions, as summarised in Table 1. The validity of the approximations involved in the theory should be tested against numerical simulations for the same model system, whereas the quality of the underlying molecular model (flexibility, binding strength, inter-molecular interactions) should be checked against experiments. The most stringent experimental test to date is the ability to reproduce the inter-colloidal potentials as measured





**Fig. 3**  $F_{\text{att}}$  as a function of distance between DNA-functionalized planes at various temperatures and coating densities, calculated via different approximate analytical theories (dashed lines) and Monte Carlo data (shaded region representing the uncertainty determined in the simulations). Different panels represent different combinations of  $\beta\Delta G^0 = [-10, -15]$  from left to right (roughly proportional to temperature), and strands grafting densities  $\rho_{\text{graft}} \approx [0.006, 0.03]$  strands per  $\text{nm}^2$  from top to bottom. For reproducibility of these results, we mention that planes have a surface area of  $14\,884\text{ nm}^2$  and either 93 or 19 (randomly grafted) strands in the high and low density regime, respectively. Color code: red = self-consistent theory with added spatial averaging of receptors,<sup>23</sup> blue = self-consistent theory, green = LCE<sub>0</sub> as in the original version,<sup>25</sup> orange = LCE<sub>1</sub>, as modified by Rogers and Manoharan.<sup>44</sup> Note that only the self-consistent theory properly describe the Monte Carlo data in all ranges of densities and  $\beta\Delta G^0$ . Both forms of LCE provide a decent comparison at high  $\Delta G^0$  (=high temperatures) and high grafting densities, but fail outside this regime.

**Table 1** Assumptions used in the theoretical work

#### Main approximations implicit in analytical theories

1. No strand-strand interaction besides those contained within  $\Delta G_{ij}^{\text{bond}}$
2. No interactions besides excluded volume between colloids and strand

#### Features included in analytical theories

1. Effects induced by changes in binding strength in solution via  $\Delta G_{ij}^0$
2. Effects of spacer configurational entropy, *i.e.* spacer structure
3. Presence of multiple bond formation
4. Competition for bonds between different strands – but only in general treatment of eqn (8) and (9) and equivalent LCE, *i.e.* effects due to competing binding configurations

in optical-tweezer experiments. Probably even more importantly than reproducing quantitative data, the presented theoretical framework also provides important qualitative insight: it allows one to rationalize various important properties of DNACCs, such as the fact the DNA-mediated binding of DNACCs has a much sharper temperature response than free DNA in solution. As a consequence, DNACCs show higher selectivity in discriminating between slightly different target DNA strands as used in bio-sensing applications.<sup>8</sup> Moreover, the theory explains why DNACCs have a higher dissociation ('melting') temperature than free DNA in solution<sup>21</sup> and it accounts for the dependence of the melting temperature of DNACCs on the type of spacer used for





grafting DNA.<sup>21,45</sup> Within the model, all these effects follow from either multivalent effects due to the large number of binding configurations, or from the entropic penalty  $\Delta G^{\text{cnf}}$  arising due to grafting constrains upon bond formation.

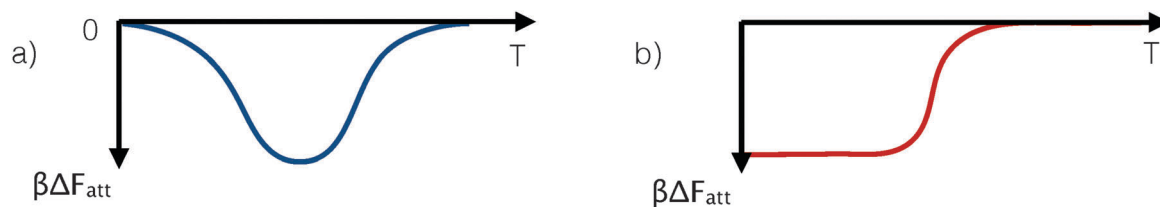
The availability of an accurate and predictive theory (eqn (8) and (9)) made it possible to explore how the competition between different strands for binding would affect the phase behaviour of DNACCs. Examples are the prediction of the design rules for mixed-strand DNACCs that can undergo re-entrant melting, and the possibility to achieve temperature independent interactions due to purely entropy-driven binding.<sup>29</sup> Although the subsequent experiments that reported both effects<sup>44</sup> used a slightly different system than the example discussed in ref. 29, the physical basis of the observed effects was the same. The easiest way to see this is by considering the predicted temperature dependence of the inter-colloidal potential of the systems of ref. 29 and 44. The underlying cause of the reentrant melting is simply that the effective attraction first increases with decreasing temperature, and then drops again, whereas the reason for a widened solid-fluid coexistence stems from purely entropy-driven (and hence temperature independent, considering  $F_{\text{att}}/k_{\text{B}}T$ ) interactions, see Fig. 4. The first behaviour is due to a thermal control of the dominant type of bond in systems featuring competing linkages,<sup>29,42</sup> while the second arises from purely combinatorial factors between configurations featuring the same number of hybridised strands, but a different number of bonds between colloids (since some of the hybridise strands do not contribute to inter-particle binding).<sup>29,44</sup>

The fact that seemingly different systems can be described by the same theory and exhibit the same behaviour, allows us to make an important general point: as long as we have quantitative information about the strength of the ligand-receptor interactions, the nature of the linkers is immaterial. ssDNA is a very convenient molecule to link different colloids together, but there are many other candidates (for example, the cucurbit[8] uril system,<sup>47</sup> or sialic acid/hemagglutinin complexes when nanocolloids/virus interactions are concerned<sup>48,49</sup>); the underlying physics remains exactly the same. We stress this fact because the development of applications non-DNA-based multivalent binding is expanding rapidly,<sup>50</sup> and such

applications can be rationalized, and designed, using the presented framework.

In the context of DNA-mediated interaction, cooperativity is often mentioned as a crucial aspect. In the theoretical description above, cooperativity plays no role; we assume explicitly that the formation of one bond does not change the binding strength of others, and yet our theoretical predictions agree very well with experiment. To avoid all confusion, we stress that, of course, the fact that one bond between two colloids has formed, makes it easier for subsequent bonds to form, simply because the loss of translational entropy of the binding partner is incurred in the first binding event only. However, that phenomenon has nothing to do with a phenomenon where, at a microscopic level, the strength of subsequent bonds is enhanced by a change in the local environment. In fact, the latter type of cooperativity is very hard to demonstrate,<sup>51–54</sup> and it would seem that all experiments where special, cooperative effects were invoked (in the context of an overly simplistic theory)<sup>51–54</sup> can, in fact, be explained within the theoretical framework that we discuss here, without cooperativity. This does not mean that there cannot be cooperative effects but simply that the evidence is lacking, if agreement with a simpler theory can be viewed as Occam's razor.

Having said this, there are definitely cases where cooperativity – but, in this case almost certainly negative cooperativity – could play a role: in nano-sized colloids, and possibly for the recently developed micron-size colloids,<sup>55</sup> grafted strands can be close enough to each other to feel the local change in the environment upon binding, be it entropic or electrostatic in origin.<sup>18,56,57</sup> Our key message is that the sharper melting behaviour of DNACCs, the key experimental finding that is often cited as evidence in support of the cooperativity model,<sup>18</sup> is perfectly well explained by our simple model that ignores cooperativity (and, in fact, by even much simpler, non-cooperative models<sup>12</sup>). On the experimental front, a sharper melting is also clearly observed for systems of micron-sized DNACCs where grafting densities were too low for any cooperativity to occur.<sup>19,20</sup> It is appealing that a simple theory that needs not invoke specific, system-dependent cooperative effects can provide a unified explanation of DNACC melting for a wide range of systems.



**Fig. 4** Schematic representation of the variation of the strength of the effective pair-attraction between DNA-coated colloids. Panel a shows the case of a non-monotonic temperature dependence of the interaction strength. Such behaviour can be realised in DNA coated colloids and results in a re-entrant phase diagram in which the solid phase melts upon both heating and cooling (see text and ref. 29 and 46). DNA-mediated reentrant melting has been observed in subsequent experiments.<sup>44</sup> In ref. 29, the DNA-mediated potentials were induced *via* inter- and intra-particle hybridisation between grafted strands, whereas in ref. 44 an indirect hybridisation *via* free-strands in solution was used. Panel b shows the case where attraction at low temperatures is entropy-dominated. In that case, the attraction strength is approximately  $\Delta F_{\text{att}} \approx -T\Delta S$ , which, when multiplied by  $\beta\Delta F_{\text{att}}$ , becomes temperature independent. A temperature-independent (scaled) interaction results in a temperature-density phase diagram with a constant width of the gas-aggregate coexistence region.



### III. Simulations

Any numerical model of DNACC systems must account for the multi-scale nature of the problem. The reason is that the physical properties of DNACCs depend on both microscopic and mesoscopic features, such as the sticky-end sequences, the structure of the spacers, and the grafting density of the DNA units, that enter into the design of the system at different length scales. In fact, the differences between various types of DNACCs are substantial. For instance, whilst nano-sized DNACCs may be coated with only a few dozen DNA strands, micron-sized DNACCs may be coated with tens of thousands of DNA strands,<sup>55,58</sup> although typically the coating density in this case is much lower than for nanocolloids.

Not surprisingly, particles that are that different require different coarse-grained modelling and simulation strategies. The modelling approaches for micron and nano-sized colloids will be reviewed in Sections IIIA and IIIC respectively. In Section IIIB we extend the discussion to micron-sized particles with mobile rather than grafted strands, *e.g.* suspensions of functionalised lipid-bilayers and emulsions. Although, in principle, fully atomistic modelling of DNACCs is conceivable, the cost of such an approach would be prohibitive and, as in other areas of colloid science, there is no need for such a brute-force approach. However, it should also be clear that microscopic details matter as the interaction between ssDNA strands depends sensitively on their nucleotide sequence. Any model, no matter how coarse-grained, must take this specificity into account.

Roughly speaking, there are three classes of coarse-grained models that account for the DNA specificity in different ways. The most detailed approaches explicitly account for the pairing of the individually-resolved nucleotides (see part 1 of Fig. 5).<sup>59–61</sup> Recently, more portable but expensive models, also capturing the double stranded structure of DNA, have been introduced.<sup>62–64</sup>

A more coarse-grained approach uses empirical rules for the binding free-energy of DNA sequences. In this description, the binding sequences are represented as structureless points (see part 2 of Fig. 5), and the binding free-energy is given by eqn (3), as discussed in Section II. Finally, a less well-founded, but nevertheless frequently used approach is based on the ‘chemical equilibrium’ approximation (see part 3 of Fig. 5). The chemical equilibrium theory treats binding reaction as a dimerisation reaction in a confined gas of sticky end groups. The advantage of this approach is that it is simple and, when the balance between free and bound ideal constructs is properly implemented (see Appendix B in ref. 23), it becomes equivalent to rigorous treatments. However these methods are practical only in the ‘mean field limit’, where the spatial distributions of sticky ends are not strongly dependent on the exact position of the grafting points.

#### A. Micron-sized particles

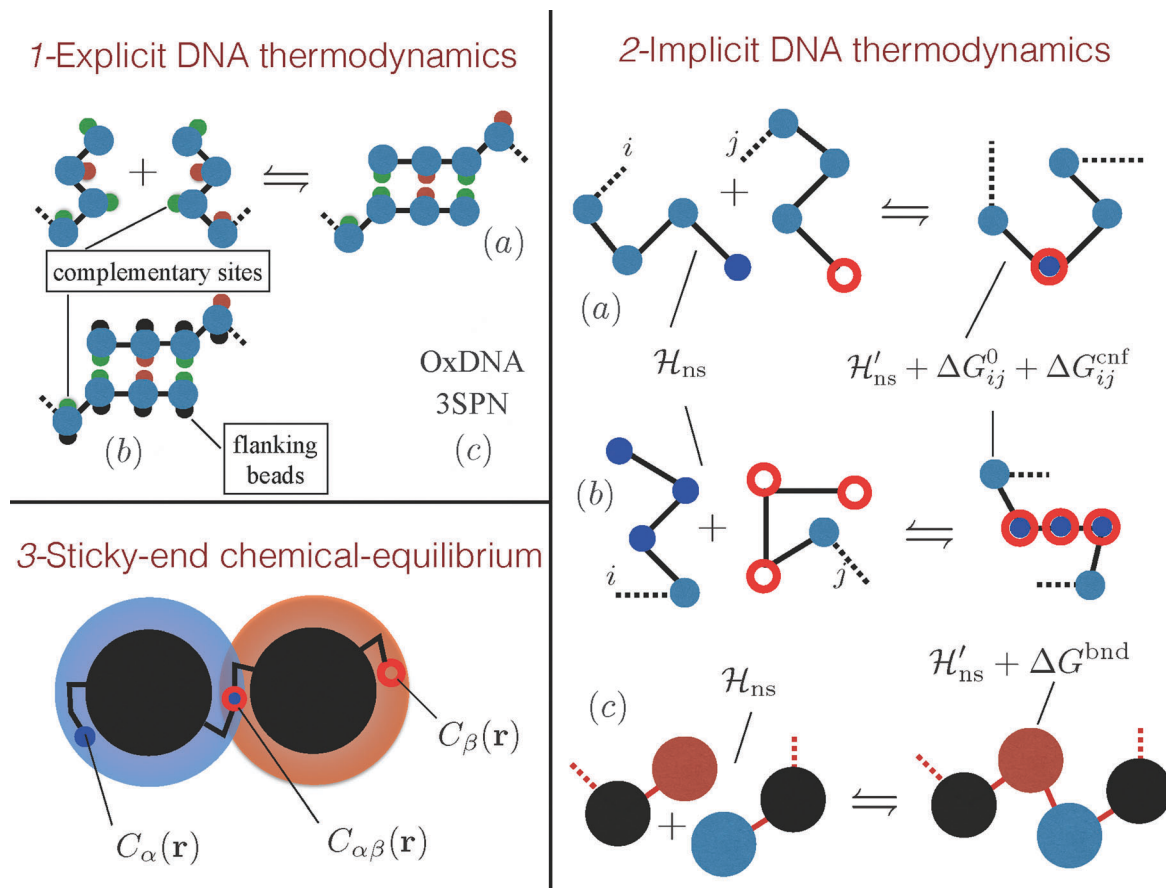
Micron-sized colloids may be coated with several thousands of constructs per particle.<sup>17,20–22,25,55,58,65,66</sup> A detailed representation of such a large number of constructs is usually not computationally feasible even when the aim is only to study the effective interaction between two colloids. The problem is,

however, not as serious as it seems, because large particles tend to have a radius of curvature that is much larger than the length of the binding strands. Under these conditions, one can use the Derjaguin approximation (see *e.g.* ref. 67) to calculate effective pair potentials between colloids<sup>26,29,68</sup> from knowledge of the distance-dependence of the interaction between two flat, parallel surfaces. If this approximation is not used, a local chemical equilibrium approach (or a “spatially averaged” version of the self-consistent equations, see ref. 23 and 27, and Fig. 3) is still tractable, but for all approaches based on the explicit calculation of interaction between complementary strands, use of the Derjaguin approximation becomes imperative. With the Derjaguin approximation, the fact that micron-sized colloids can be coated with very large numbers of DNA strands does not create a technical challenge. As discussed Section II, most theoretical approaches developed to study DNACCs are based on the assumption that non-binding strand–strand interactions can be ignored. We stress that this assumption is not essential (and questionable for nano-sized colloids: see ref. 31 and 32 and Section IIIC), but in many cases, the neglect of non-binding strand–strand interactions is justified, in which case an accurate estimate of  $\Delta G^{\text{cnf}}$  is all that is required to obtain quantitatively accurate potentials from computer simulations. In practice, we calculate the partition function of unbound strands and the configurational part of the binding free energy of two strands,  $\Delta G^{\text{cnf}}$ , by simulation. The strand–strand binding free energy  $\Delta G^0$  is not obtained from simulation but from the empirical SantaLucia rules.<sup>69,70</sup> Again, more sophisticated rules could be used where necessary. Once these two ingredients are known, and if strand–strand interactions can be neglected, the theoretical approaches described in Section II can be used to calculate accurate inter-particle potentials. As was demonstrated in experiments of Bracha for DNA-chips *et al.*<sup>71</sup> strand conformations are weakly affected by DNA–DNA interaction up to thousand strands per  $\mu\text{m}^2$ . Moreover, a rough estimate<sup>26</sup> suggests that the contribution of non-bonded spacer interactions to the free energy of the system is of the order of one  $k_{\text{B}}T$  unit per strand, which is much smaller in absolute terms than the typical values of the configurational and combinatorial free-energy that are around  $\approx 10k_{\text{B}}T$  for typical spacers.<sup>21</sup>

A notable exception of a micron-size system where non-bonded strand interactions may be significant is provided by the densely grafted micron-sized colloids studied by Pine and co-workers.<sup>55,58</sup> The reason why higher grafting densities could be reached in ref. 55 and 58 is that in these experiments, grafting was achieved by using click chemistry, rather than *via* the use of rather bulky streptavidin constructs that anchor the biotinylated spacers.

Having sketched the various theoretical approaches to predict the binding strength of DNACCs, we next consider how to estimate the change in the configurational free energy  $\Delta G^{\text{cnf}}$  of the grafted strands as the distance between colloids is varied. In general,  $\Delta G^{\text{cnf}}$  can be computed using Monte Carlo simulations.<sup>23,24,72</sup> Practical aspects of these calculations for ssDNA constructs are discussed in Appendix B. In this modelling, it is important to account for intra-chain Coulomb interactions.





**Fig. 5** Modelling DNA hybridisation in DNACC systems. (1) Models that feature explicit DNA thermodynamics attempt to recover DNA hybridisation with an explicit modelling of the dimerisation process. (a) Ref. 59 and 60 used a bead-and-spring ssDNA representation decorated by sites that selectively attract. (b) This model was ameliorated by ref. 61 that introduced flanking beads to avoid binding between more than two bases. Although the molecular mechanism for binding is retained in these models, none of them tries to match the real thermodynamics of DNA hybridisation but simply use an effective nucleotide–nucleotide interaction which should be considered as a tunable model parameter. (c) Nucleotide level models, such as SNP3<sup>64</sup> or OxDNA,<sup>62,63</sup> have been parametrised using thermodynamic experiments. (2) Models that feature an implicit DNA hybridisation rely on empirical estimates of the hybridisation free energy of the sticky ends ( $\Delta G_{ij}^0$ ).<sup>69,70</sup> (a) Chains, interacting by non selective Hamiltonians ( $\mathcal{H}_{ns}$ ), hybridise by coalescing the two reactive end-elements according to  $\Delta G_{ij}^0$  and to a configurational cost ( $\Delta G_{ij}^{cnf}$ ), that depends on the algorithm that generates the hybridised chain.<sup>23,24,31,72</sup> Notice that the non selective Hamiltonians of the free and bound states are not the same.<sup>23,31,72</sup> (b) A similar scheme was previously proposed by Mladek *et al.* that also considered the case in which more reactive sites bind resulting in a rigid dsDNA segment.<sup>31,32</sup> (c) For long  $\lambda$ -DNA ref. 74–76 used a blob representation in which reactive blobs are chained according to an hybridisation free energy  $\Delta G^{bnd}$ . (3) For ideal spacers<sup>23</sup> a local chemical equilibrium balance can be used to calculate the fraction of hybridised strands ( $x_\beta$ ).<sup>23,25,28</sup> The concentrations of the sticky ends around the colloids  $C(r)$  are calculated by the end-to-end distribution functions of the tethered spacers,<sup>25</sup> while the equilibrium constant between sticky ends is  $\Delta G_{ij}^0$ .

For instance, ssDNA constructs have been modelled as uniformly charged, freely-jointed chains (FJC)<sup>34</sup> with screened Debye–Hückel interaction (see, *e.g.* ref. 73). Such an approach has been used to model micron-sized<sup>23,24,72</sup> and, as we will discuss later, nano-sized<sup>31,32</sup> DNACCs. When accounting for electrostatic interactions in this way, the electrostatic contribution has to be removed from the phenomenological DNA hybridisation free energies, to avoid double counting.

The numerical approach to compute the configurational contribution to the DNACC interaction described in Appendix B is efficient for short to medium length linkers that can be coarse-grained into at most a few dozen Kuhn segments. However, the same methods would be computationally unfeasible when applied to very long chains. Numerical studies of colloids functionalised by very long DNA strands such as  $\lambda$ -DNA<sup>77,78</sup>

require a further simplification of the description of the DNA degrees of freedom. Bozorgui and Martinez<sup>74–76</sup> employed a numerical approach based on the ‘blob’ representation of long polymer chains<sup>79,80</sup> in which entire DNA constructs are lumped into sites interacting *via* gaussian potentials.<sup>79–81</sup> Multi-blob models have been used by Mladek *et al.*,<sup>31,32</sup> but the approach, whilst fairly accurate, quickly becomes cumbersome. The potentials between the blob and the colloids have been parametrised by using explicit ‘all-monomer’ simulations, while the hybridisation reactions between complementary strands is simulated using a dedicated MC move that attempts to join two complementary blobs with a cost equal to  $\Delta G_{ij}^{bnd}$ , where  $\Delta G^{cnf}$  is given by the energy of an harmonic spring connecting the centres of the two blobs of the form  $\beta \Delta G^{cnf} = 0.534(r/R_g - 0.730)^2$ , where  $R_g$  is the gyration radius of the strand and the numerical values



have also been parametrised using all monomer simulations (see part 2c of Fig. 5).

The ultimate aim of coarse-grained modelling of DNACCs is to predict their phase behaviour. Once the tools are in place to compute the DNACC interactions, standard free-energy calculations<sup>36</sup> can be used to compute phase diagrams. Examples of such studies can be found in ref. 74–76. These simulations showed that DNACCs exhibit very interesting phase behaviour. For example, the simulations showed that the liquid–crystal–vapour triple point disappears as the number of grafted chains drops below a threshold of around 12. In that case, the ‘ground-state’ at zero pressure but high densities is a liquid, not a crystal.<sup>76</sup> Recent simulations on ‘patchy’ colloids,<sup>82</sup> showed similar behaviour in such systems.

## B. Mobile tethers

Recent experiments have studied DNACCs where the DNA strands are mobile, rather than grafted in fixed positions. Examples include emulsions<sup>83–85</sup> and lipid bilayers<sup>33,86,87</sup> functionalised by hydrophobised DNA constructs terminated by reactive sticky end sequences.

It is straightforward to extend the theoretical tools to model DNACCs with grafted DNA strands to the case where these strands are mobile: due to the translational degrees of freedom of the binders, the hybridisation free-energy needs to include an extra configurational term ( $\Delta G^{\text{trn}}$ ) that accounts for the constraint on the relative position of two tethering points when bound

$$\Delta G_{ij}^{\text{bond}} = \Delta G_{ij}^0 + \Delta G_{ij}^{\text{cnf}} + \Delta G_{ij}^{\text{trn}}. \quad (18)$$

This confining entropic cost  $\Delta G_{ij}^{\text{trn}}$  destabilises DNA pairing. However, the combinatorial gain is higher in system of mobile tethers because each binder can potentially bind more complementary partners than in a system with fixed tethering points. A quantitative analysis of all the entropic and configurational terms of eqn (18) for a system containing both bridges (*i.e.* bonds between strands residing on different particles) and loops (*i.e.* between strands on the same particle, when this bears complementary sequences) has been given in ref. 33. Interestingly, in these systems the self-consistent theory developed in ref. 23, 27 and 42 (eqn (8) and (9)) becomes exact. This was demonstrated in ref. 28, starting from an exact free energy functional  $\mathcal{F}(\{n\}, \{\Delta G^{\text{bond}}\})$  calculated as a function of the number of bridges between bound colloids ( $\{n\}$ ) and the matrix of the hybridisation free energies  $\Delta G_{ij}^{\text{bond}}$  (where  $i$  and  $j$  now represents a strand type, not a single specific strand). Taking the large number of binder per particle limit, a saddle-point approximation provided the most probable number of bridges between colloids ( $\bar{n}$ )

$$\left. \frac{\delta \mathcal{F}(\{n\}, \{\Delta G^{\text{bond}}\})}{\delta n} \right|_{n=\bar{n}} = 0. \quad (19)$$

Remarkably eqn (19) is equivalent to eqn (9). Using  $\{\bar{n}\}$  one can calculate the attractive potential between colloids

$$F_{\text{att}} = \mathcal{F}(\{\bar{n}\}, \{\Delta G^{\text{bond}}\}) \quad (20)$$

and recover exactly the expression given in (8) (once the number of bridges  $\bar{n}$  are expressed in term of probability of making

bridges between colloids). It is possible to generalise this discussion to systems featuring different types of bonds including not only bridges but also loops.<sup>33</sup> Slightly different expressions are obtained in presence of infinite reservoirs of binders as shown in the study of a vesicle interacting with a large supported lipid bilayer.<sup>87</sup>

Eqn (19) and (20) were used to simulate rigid colloids functionalised by mobile strands (a possible experimental realisation being that of ref. 86) in which colloidal suspensions are sampled using an effective interaction free energy  $F_{\text{att}}(\{\mathbf{R}_i(t)\}) + F_{\text{rep}}(\{\mathbf{R}_i(t)\})$  that only depends on the positions of the colloids  $\{\mathbf{R}_i(t)\}$ .

We note that the behaviour of DNACCs with mobile linkers can be qualitatively different from that of DNACCs with grafted linkers, in other words: mobile linkers introduces new physics in the system. For example, ref. 28 showed that the inter-colloidal potential induced by DNA-hybridisation of mobile linkers is intrinsically a multi-body interaction. This is due to the fact that each construct on a given particle can bridge all the neighbouring particles. Hence, different neighbouring colloids may compete for the same linkers. As the number of bonds with different neighbours are now correlated, the free-energy cannot be decomposed into the sum of pair-interactions. The fact that DNACCs with mobile linkers have intrinsically multi-body interactions can be exploited used to control the ‘valency’ (preferred number of complementary neighbours) of a particle.<sup>28</sup> This is interesting because conventional approaches to tune the valency of colloids are based on the use of different number of interacting ‘patches’. However, the preparation of well-controlled patchy colloids is non-trivial. In contrast, DNACCs with mobile linkers can have a preferred valency without the need of physically creating real patches on the surface of the colloid.<sup>41</sup>

Even more complex effects should be expected for systems of mobile strands where the colloid to which they are attached is deformable, as in the case of DNA-functionalised vesicles,<sup>33</sup> where the flexibility of the lipid membrane plays a non-negligible role<sup>87</sup> in determining the bond-mediated interaction. For instance, Hu and collaborators<sup>88</sup> considered the interaction between two membranes that were coated with complementary proteins (rather than DNA). They found that the equilibrium constant for protein–protein binding (*i.e.*  $\Delta G^{\text{bond}}$  in our nomenclature) decreases by a factor that is controlled by the roughness of the membranes.<sup>88</sup> This effect, when translated to DNA-mediated interaction, can be expected to have important consequences for the self-assembly.

## C. Nano-sized particles

Modelling the interaction between nano-size DNACCs is more demanding than modelling their micron-sized counterparts. The underlying reason is simple: the length-scale separation that justifies the factorisation between the different driving forces and facilitates modelling of micron-sized particles is less pronounced for nano-sized particles.

There are other differences as well: many of the earlier experiments on DNACCs (although not the most recent ones<sup>55,58</sup>) used a bulky streptavidin–biotin construct to tether DNA strands to the surface of micron-sized colloids. In contrast, DNA anchoring to





gold nano-particles is usually achieved by bonding the (much smaller) thiol group to the gold surface. As a consequence, the density of the DNA strands on nano-sized colloids is sufficiently high that the interaction between different DNA-strands cannot be neglected for a quantitative modelling. Constructing a coarse-grained model of nano-sized DNACCs is therefore much more system-specific than for micron-sized colloids. A case in point is the coarse-grained model developed by Mladek *et al.* in an attempt to reproduce a typical experimental system.<sup>5</sup> Construction of the model required three coarse-graining steps. In the first step, simulations were performed on nano-colloids (12 nm diameter) functionalised by 60 homogeneously charged freely-jointed chains representing the 665 base-pair constructs used in experiments.<sup>5</sup> At this level of description, no DNA hybridisation was taken into account. Hence, the particles were simply treated as nano-colloids sterically stabilised with densely grafted chains. The simulations probed the repulsive interaction between two such colloids. Subsequently, a multi-blob model was constructed to reproduce the computed repulsive interaction between the nano-colloids. The ends of the coarse-grained chains were then functionalised with units that could hybridise. However, for the nano-sized colloids, the spatial extent of the 'sticky' end group could not be ignored, nor the fact that, upon hybridisation, two such groups combine to form a double helix that is more rigid than the unbound sticky ends. Hence, although in this model too the hybridisation free energy is calculated using the SantaLucia nearest neighbour rules,<sup>69,70</sup> the conformational changes upon hybridisation are accounted for in the model (see part 2b of Fig. 5).

Using this fairly complex model, the authors computed the crystallization behaviour of the nano-sized DNACCs and they computed the structural properties of the crystalline solid. Encouragingly, the simulations predicted the correct stable crystal structure and gave very good estimates for the temperature at which the solid melts. However, the estimated crystal lattice spacing was off by some 10% (presumably because the experimental estimates of the persistence length of ssDNA show a considerable spread). The other important finding was that three-body interactions play an important role and must be correctly included in order to predict the structure and melting of the crystalline phase. As the above discussion shows, coarse-grained modelling of nano-sized DNACCs is not simple. However, with the use of a coarse-grained polymer model and the empirical SantaLucia nearest neighbour rules, it is still much more efficient than a brute force simulation of nano-size DNACCs. Indeed, both the system sizes (hundreds to thousands of colloids) and the timescale for hybridisation are such that atomistic simulations of the phase behaviour of DNACCs is out of the question for the foreseeable future, whilst calculation of the pair interaction between nano-sized DNACCs is, at least at present, not an attractive proposition.

There are, however, good reasons to use models for DNA that, whilst not atomistic, do reproduce the structure and dynamics at the nucleotide level. In fact, much effort has been devoted to the development of coarse-grained models that are capable of reproducing the hybridisation free energy, the binding kinetics and

the mechanical properties of DNA complexes.<sup>62–64</sup> As the interactions that drive DNA hybridisation are primarily the hydrophobic forces between neighbouring bases, such models require a nucleotide-level description of the biopolymer.

Although there have been extensive and very successful simulations of systems with few DNA strands, the models used are still not an attractive proposition to simulate the collective behaviour of DNACCs, as such simulations would require the study of systems of hundreds (or thousands) of colloids each coated with dozens of strands. Moreover, computing the thermodynamic properties from equilibrium simulations requires sampling times that are much longer than the time for a typical hybridisation/de-hybridisation events. Hence, at present, this is not yet possible with the models of ref. 62–64.

For this reason, even more coarse-grained – but still nucleotide-based – descriptions have been developed starting with the work of Starr and Sciortino.<sup>59,60</sup> In the approach of ref. 59 and 60, the nucleotides are described as (repulsive) spheres functionalised with a binding site. Neighbouring monomers are bound together by a so-called FENE potential.<sup>89</sup> A bending rigidity term between each three consecutive monomers is also used. A binding site, specifying the type of base, is connected to each monomer *via* a FENE potential. These binding sites selectively interact with complementary sites *via* Lennard Jones potentials. This model has been used to study the assembly of dendrimer-like structures functionalised by four DNA arms,<sup>59,60</sup> self-assembly mediated by linkers<sup>90</sup> and crystallisation of nanoparticles.<sup>91</sup> A model that is closely related to that of Starr and Sciortino<sup>59</sup> was deployed by Knorowski and collaborators<sup>92</sup> (see part 1b of Fig. 5). As in ref. 59 and 60, DNA constructs were modelled by mean of beads carrying selective binding sites. However, two flanking repulsive sites were added to every bead belonging to the sticky end part of the construct. The function of the flanking beads is to provide directionality to the interaction between complementary bases and to prevent (unphysical) multi-base interaction. This model has been used to probe the stability of different crystals<sup>92,93</sup> and to investigate the dynamics of crystallisation.<sup>61</sup> Recently the model of Starr and Sciortino<sup>59</sup> has been used by Theodorakis and collaborators to study the pair interactions between nanoparticles functionalised by up to eight constructs.<sup>94</sup> This study highlighted the need for a potential further refinement of the model, as the simulations showed the presence of spurious bound states in which the ssDNA strands that form a duplex were oriented parallel, rather than anti-parallel. Parallel hybridisation is not found in real DNA, as the strongly polarity-dependent Watson–Crick pairing between complementary nucleotides uniquely favours anti-parallel hybridised sequences.

The model of Knorowski has been further developed by Li *et al.*<sup>95,96</sup> In particular the size of the beads in the sticky part and in the spacer part of the construct have been differentiated to qualitatively account for the different mechanical properties of the ssDNA and of the dsDNA spacers. This model is parametrised in such a way that each bead corresponds to  $\approx 2$ –3 bases. This relatively detailed model was subsequently used to parametrise an effective pair-potential that is composed of a short range repulsion due to the electrostatics of the system, and an attractive



interaction provided by the hybridisation of the sticky-ends.<sup>97,98</sup> Using the effective model the authors rationalised experimental finding on the dynamics of crystallisation,<sup>97,99</sup> while the detailed model has been used to probe the stability of the crystals<sup>95,96</sup> and to investigate the microscopic structure of the facets of DNACCs-made crystal.<sup>99</sup> Detailed molecular models have also been used to investigate interaction between DNA constructs and the nanoparticle substrate (*e.g.* ref. 100 and 101), and to study the changes in the structure of dsDNA when used to assemble crystalline structures.<sup>102</sup> However, these studies go beyond the scope of the present review, where we focus on the physical properties of DNACCs that are related to the reversible binding of multiple DNA complementary pairs, rather than the formation of a single pair formation.

#### D. Modelling perspectives

There are, in principle, two ways to construct a proper coarse-grained model for DNACCs. One (the top-down) approach is to derive a coarse-grained model from a fully atomistic model for DNA,<sup>62–64,104</sup> and use it to simulate DNACCs. This type of approach thus starts with a system at the top of the “complexity scale”, and uses systematic coarse-graining techniques to move towards the bottom of the same scale. Such an approach still retains a molecular based description, and should be able to describe the hybridisation thermodynamics of DNA. The second approach is instead to start with the choice of the ingredients of the coarse-grained model (*e.g.* ideal chains, rods, monomers, blobs, hybridisation sites *etc.*) and then to use comparison with the available experimental information to tune its parameters.

At present, the top-down approach is not yet as extensively used as that comparing coarse-grained models directly with experiment. An encouraging example in this direction is the work of Lequieu *et al.*<sup>105</sup> and Ding *et al.*,<sup>106</sup> who recently reported studies of pair potentials between nanoparticles functionalised by few constructs of reactive DNA, using the 3SPN model previously used to study the dimerisation of DNA in solution (ref. 64 and 104). The OxDNA model of Ouldridge and coworkers<sup>62,63</sup> is another valuable CG model that has been tested in multiple DNA nanotechnology systems. To our knowledge, OxDNA has not yet been used to improve the modelling of DNACCs.

There are other aspects of the modelling that need to be improved. One is related to the naive use of nearest-neighbour rules to compute DNA binding free energies in constructs. Recent experimental/theoretical investigation by Di Michele *et al.*<sup>103</sup> showed that the SantaLucia nearest-neighbour rules<sup>69,70</sup> overestimate the hybridisation free energy  $\Delta G^0$  in the case of DNA oligomers that are extended to include strings of non-complementary bases.<sup>103</sup> Considering for instance two complementary sequences of ssDNA decorated by two inert strings at their 5' terminals (snapshot of Fig. 6a), ref. 103 reported that  $\Delta G^0$  increases with the length of the inert tails until reaching a plateau for a number of inert bases equal to  $\approx 5$ –7 (plot in Fig. 6a). In particular the presence of the inert tails destabilises double helix formation and hence decreases the melting temperature of the oligomers. Such an effect is due to the electrostatic interactions between the charged backbones of the DNA that is not included in the nearest neighbour rules.<sup>107</sup> Quite surprisingly ref. 103 showed that the shift in the hybridisation



**Fig. 6** (a) Shift in the hybridisation free energy as due to the electrostatic interactions between the inert tails (blue beads in the top rendering of the two oligomers) and the paired bases as a function of the number of inert bases.<sup>103</sup> Electrostatic interactions are modelled using the linearised Debye–Hückel<sup>73</sup> theory at three different salt concentrations. Further details of the simulation method are reported in ref. 103. (b) SantaLucia hybridisation free energy of two oligomers (3'-AACCGACAG-5' and complementary strand) with and without two (G) dangling bases attached to the 5' terminals of the reactive sequences for 100 mM NaCl concentration.<sup>69,70</sup> The stacking contribution of the SantaLucia rules (blue arrow in the figure) is comparable with the tail contribution (see part a). Notice that an error of  $\pm 1 k_B T$  in the nearest-neighbour rules<sup>69,70</sup> results in a bigger uncertainty of the melting temperature (black arrows).

free energy due to the tails (at salt concentration that are typical of DNA coated colloids experiments<sup>25</sup>) can be as large as the stacking contribution of the nearest-neighbour rules due to the first dangling (*i.e.* non-hybridised) base in the nucleotide sequence, a standard term entering the nearest-neighbour rules (see Fig. 6b). The relevance of the inert tails to the calculation of the hybridisation free energy  $\Delta G^0$  seems to limit the degree of 'responsible' coarse graining of DNA mediated interactions (also in other DNA nanotechnology systems<sup>108</sup>). In particular if we consider models with explicit hybridisation thermodynamics, the importance of the electrostatic interactions between the inert tails and the DNA duplex requires an explicit description of the charged backbones in the tails and in the duplex at the atomistic level. Interestingly the latest version of OxDNA introduced Debye-Hückel interactions<sup>73</sup> between the backbone sites.<sup>63</sup>

The problem is that the binding free energy computed for a typical configurations of the dangling segments of the free oligomers,<sup>103</sup> may not be transferable. This is the case, for instance, if the tails are more stretched when bridging two colloids than when they are free in solution. Hence, estimates of  $\Delta G^0$  obtained from experiments for free strands might not be reliable once the same strands have been grafted. These 'minor' effects are not so minor when one realises that a bias of one  $k_B T$  in  $\Delta G^0$  results in a shift in the melting temperatures of about  $\approx \pm 3.5$  K (see Fig. 6b). Hence, when analysing accurate experiments,<sup>25</sup> neglect of the dangling tail effects can affect the interpretation of the results, once the experimental resolution gets better than a few Kelvin.<sup>17</sup>

There is also the 'no-free-lunch' theorem of coarse graining. As was pointed out in a different context many years ago by Louis, the design of an optimal model that carefully reproduces the structural properties of a system may result in a model that is not representative of the thermodynamics of that system (and *vice versa*).<sup>109</sup>

Finally we stress something that has been stressed many times: structures may be mechanically stable, but thermodynamically unstable. Demonstrating that one structure is more stable than another typically requires the calculation of the free energy (or, more precisely, the chemical potential) of both systems. Using MC or molecular dynamics to see what kind of structure forms by annealing from a disordered state<sup>92,93,95-97</sup> is interesting, but it does not demonstrate that the structure that forms is stable (in fact, Ostwald's rule<sup>110</sup> suggests the opposite). The problem is that the first stages of the self-assembly process are those with a smaller barrier from the starting disordered state. These states can simply represent mechanically stable metastable states, which would transform to the true thermodynamically stable state at much longer times (if at all). These longer times may not be accessible with experiments, and are certainly many orders of magnitude larger than the time-scales accessible by computer simulations. We also notice that, in order to correctly reproduce the stable phases observed in experiments, studies using this simulated annealing approach typically start with simulations boxes compatible with the shape and number of particle experimentally observed, which must thus be known *a priori*. Since maximal packing appears to be, at least for conventional binary DNACCs system, one of the major driving force behind formation

of specific lattices,<sup>16</sup> starting with experimentally available conditions necessarily biases the simulation toward the expected structure. In short: constructing a phase diagram requires the calculation of the free energy of each possible phase,<sup>36</sup> unless a direct, hysteresis-free transition between the two phases can be observed. One of the main challenges in the simulation of the phase diagram of DNACCs was precisely the need to carry out free-energy calculations<sup>31,32,75,76</sup> that properly account for the hybridisation between the binders. Fortunately, efficient MC algorithms have been developed to calculate the density of states of an aggregate as a function of the number of bridges between particles, whose integration can then be used to calculate the free energy of the crystal.<sup>72,75,76</sup> These algorithms have been successfully combined with various techniques to select among the many possible candidate phases. For example, in view of the rich polymorphism offered by DNACCs, genetic algorithms<sup>111,112</sup> have been used to identify possible crystal structures, whose relative thermodynamic stability was then probed by free-energy calculations, providing phase diagrams in quantitative agreement with experiments without an *a priori* knowledge of crystals parameters.<sup>31,32</sup>

## IV. Conclusions

Theoretical and computational modelling have proven to be powerful tools to rationalise the observed behaviour of DNACCs,<sup>17,21-23,25,92,95-97,113</sup> and, in some cases, to predict novel phenomena,<sup>29,44</sup> with some theoretical predictions for unconventional yet experimentally accessible systems still waiting to be tested.<sup>28,76</sup>

Understanding DNACCs behaviour means understanding their mutual interactions as a function of the system's parameters. In this regard, we have shown how the presently available theoretical framework can be used to design interactions of almost arbitrary complexity: how these could lead to interesting effects (see for instance ref. 114 and 115 for some applications) is yet to be systematically investigated. An important factor to highlight in this regard is that the analytical framework we described assumes thermodynamic equilibrium for bond formation, *i.e.* it implicitly considers the bonding network to be equilibrated on the timescale of colloidal diffusion. Hence, in order to connect the effective inter-particle potential predicted by theory with experimental observations (that are of finite duration) one requires both the bond-formation and bond-breaking rates to be fast compared to the timescale for colloidal diffusion. Until now, this was a strong assumption. However, although this was not true for general bonding schemes previously typically employed, Rogers and Manoharan<sup>44</sup> have recently shown a very powerful and general scheme exploiting toe-holding mechanisms that can make such assumption essentially correct. Proof of this point is already observable in their experimental results, whose trends could be rationalised using a local chemical equilibrium approach that implicitly makes such an assumption. Hence, we would expect such scheme to lead to a more reliable comparison between theory and experiments, opening new and truly exciting perspectives for this field.



Most of the focus in the DNACCs has probably been on their self-assembly property, and much remains to be done to exploit this system to its full potential, in particular in the design of functional materials for applications. In our view, however, other DNACCs based technologies will benefit from interpreting experiments in view of the presented framework, selective drug delivery and biosensing being two likely candidates. It should be stressed that, besides describing DNACCs – systems that have important nanomedicine applications in their own regard<sup>8–10</sup> – the theoretical and computational methods outlined here apply more generally to all systems whose interactions are dominated by the formation of non-covalent, multiple and reversible ligand–receptor bonds, *i.e.* supramolecular, multivalent based systems. A wide variety of such systems are already under intense investigation for nanomedical applications,<sup>48,49,116,117</sup> and we envisage important cross-fertilization of ideas between these fields and DNACC research. In particular, understanding how to embed binding selectivity in multivalent-based systems, and not just increase their binding strength, seems to be a crucial point where theoretical work<sup>118</sup> might prove useful to rationalise observed effects,<sup>116</sup> and speed up the rational design of applications.

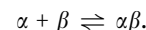
Finally, we stress that there is much need for further theoretical and computational developments. For applications that will require quantitative predictions, more accurate experiments are needed to inform the parametrisation of existing, or the development of, novel models. An illustration is the tail effect discussed in Section III. Understanding the role of non-specific attractive interactions, and also of strand–strand interactions, are further challenges that have still not been properly investigated. The need for such improved modelling has become urgent with the arrival of recent experiments of the Pine group, who was able to reach coating densities that are much higher than what was previously achievable.<sup>55,58</sup> Interestingly these experiments reported enhanced crystallisation rates; it would be interesting to understand these results using properly extended models. Finally, other challenges clearly arise when considering systems where the colloid itself can be deformed upon binding (hence effectively coupling  $\Delta G^{\text{cnf}}$  with the shape of the colloidal core), such as the case of functionalised vesicles.

After the long gestation period that followed the early work by Mirkin and Alivisatos, the field is moving rapidly. Exciting times lie ahead of the DNACCs field. We hope that this review will have clarified some of the key unifying concepts that need to be understood in the move towards the rational design of DNACC-based structures. In this regard, we make freely available a python-based set of routines that we developed to calculate ligand–receptor mediated interaction free energies using our self-consistent equations at the following address <https://github.com/sangiole/DNACC>.

## Appendix A: Breakdown of local chemical equilibrium and mean-field approximations

To illustrate the breakdown of LCE and mean-field like approaches in the limit of small numbers of binding strands,

first consider the normal dimerisation reaction between two “molecules”  $\alpha$  and  $\beta$ .



The condition for chemical equilibrium is

$$\mu_\alpha + \mu_\beta = \mu_{\alpha\beta} \quad (\text{A1})$$

We assume that the solutions are ideal. Then the partition function of species  $x$  ( $x = \alpha, \beta$  or  $\alpha\beta$ ) is

$$Q_x = \frac{V^N q_{\text{int}}^N(x)}{N! \Lambda_x^{3N}}$$

where  $q_{\text{int}}(x)$  is the contribution to the partition of  $x$  due to all internal degrees of freedom. The equilibrium condition A1 can be written as

$$\left( \frac{C_{\alpha\beta}}{C_\alpha C_\beta} \right) = \frac{q_{\text{int}}(\alpha\beta)/\Lambda_{\alpha\beta}^3}{q_{\text{int}}(\alpha)q_{\text{int}}(\beta)/(\Lambda_\alpha^3 \Lambda_\beta^3)} \quad (\text{A2})$$

where  $C_x$  is the number density of species  $x$ . To compare with experiments, it is conventional to express the equilibrium condition such that the concentrations can be expressed in mol per liter.

$$\frac{[\alpha\beta]}{[\alpha][\beta]} = K$$

In order to convert to number densities, we note that  $[x] = C_x/\rho^0$  where  $\rho^0$  is the “standard” number density (1 molar =  $6.022 \times 10^{26}$  molecules per  $\text{m}^3$ ). Then

$$\left( \frac{C_{\alpha\beta}}{C_\alpha C_\beta} \right) = K/\rho^0 = \exp(-\beta \Delta G_{\alpha\beta}^0)/\rho^0$$

Next consider the case where we still have an ideal gas mixture, but now with only very few molecules in the ‘reaction volume’ (in the case of tethers, this translates into: ‘a tether of a given type (say  $\alpha$ ) can only bind with a small number of nearby complementary tethers. In that case, LCE breaks down. To see this, consider the extreme (but not unrealistic) case of a volume  $V$  containing a single molecule of type  $\alpha$  and a single copy of  $\beta$ . As before,  $\alpha$  and  $\beta$  can react to form  $\alpha\beta$ . Let us denote by  $P$  the probability that molecule  $\alpha$  (or  $\beta$ ) is part of a dimer, then  $P/(1 - P)$  is simply given by the ratio of the corresponding single-particle partition functions:

$$\frac{P}{1 - P} = \frac{q_{\text{int}}(\alpha\beta)/\Lambda_{\alpha\beta}^3}{V q_{\text{int}}(\alpha)q_{\text{int}}(\beta)/(\Lambda_\alpha^3 \Lambda_\beta^3)} = K/(V\rho^0)$$

Now compare this with the LCE expression. The number density of  $\alpha(\beta)$  equals  $P/V$  and the number density of  $\alpha\beta$  is  $(1 - P)/V$ . Hence, LCE predicts:

$$\frac{P}{(1 - P)^2} = K/(V\rho^0),$$

which is clearly not correct. Hence, we should not expect LCE and in general mean field theories not accounting for the exact position of the tethering points) to work if only small numbers of monomers can interact. However, this is precisely the situation if the grafting distance between strands is not





small compared to the effective length of the strands. On the other hand Fig. 3 shows that the self-consistent approach performs very well also at low grafting density.

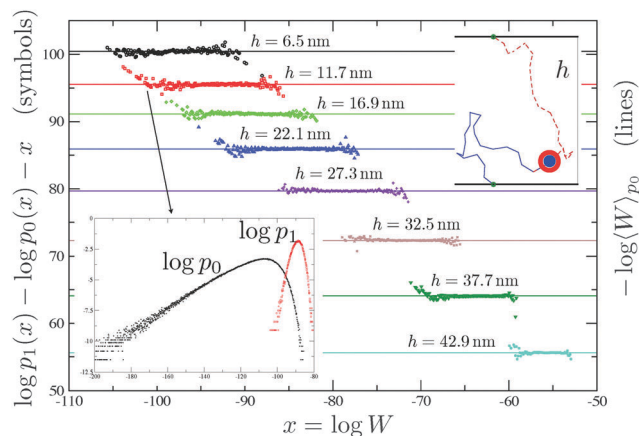
## Appendix B: Numerical estimate of configuration contribution to binding free energy

In this appendix, we describe how the hybridisation free energy between two ssDNA constructs can be calculated numerically. The method was employed in ref. 23 and 24.

Two independent Monte Carlo runs are used to generate two interacting FJCs with free and bound end points, respectively, using the Rosenbluth algorithm.<sup>36</sup> In the bound case one grows a single longer chain using a bias to constraint the two end-points to the tethering points<sup>119,120</sup> (see the top inset of Fig. 7). In ref. 23 and 24, while growing a hybridised chain, we biased the choice of the new segments using the end-to-end distribution function of an ideal FJC of length equal to the number of missing segments<sup>119</sup> (notice that this is not the unique choice possible<sup>120</sup>). By sampling the corresponding Rosenbluth weights<sup>36</sup> (see the right y-axis of Fig. 7 for the bound case) it is possible to calculate  $W_{ij}$  and  $W_j$

$$W_{ij} = \langle W^{(f)} \rangle \quad W_j = \langle W^{(b)} \rangle \quad (\text{B1})$$

(where  $W^{(f/b)}$  are the Rosenbluth weights calculated for free and bound constructs respectively) that can be used in eqn (5) to calculate  $\Delta G^{\text{cnf}}$ . Notice that in this case  $p_{\text{ee}}$  is the end-to-end distribution function of an ideal FJC with the number of segments



**Fig. 7** Inset bottom: The algorithm used in ref. 23 and 24 generates chains with Rosenbluth weights that are distributed according to a distribution function ( $p_0$ ) which is different from the distribution obtained using equilibrium runs ( $p_1$ ). The average of the Rosenbluth weight using  $p_0$  (right y-axis), together with the corresponding result for free constructs, is used in eqn (5) to calculate  $\Delta G^{\text{cnf}}$ . Using the overlapping between  $p_1$  and  $p_0$  and the results of ref. 121 we can calculate, in an independent way,  $-\log(W)p_0$  (left y-axis and symbols). The agreement between symbols and the straight line is perfect. The algorithm of ref. 72 samples between free and bound states in a way that the Rosenbluth weights are distributed with  $p_1$ . In this case  $\Delta G^{\text{cnf}}$  can be measured comparing the time spent by the system in the free and in the bound state.<sup>72</sup>

equal to the hybridised construct (see top inset of Fig. 7) and that  $W^{(f/b)}$  also account for the non-selective interactions between constructs. In the particular case of non-interacting constructs, a further simplification can be used since  $W_{ij}$  factorises into the product of two independent Rosenbluth weight ( $W_{ij} = W_i W_j$ ), and one can thus use eqn (5) instead of the more general eqn (4). Notice that for interacting strands ( $W_{ij} \neq W_i W_j$ ) a consistent approach would require to include the interactions between constructs also when calculating the repulsive part of the potential (see first term of eqn (1)).

The accuracy of the aforementioned calculations depends on the quality of the sampling, which can be probed using the overlapping method developed in ref. 121, a typical output being shown in Fig. 7 (see the caption for further details). Moreover ref. 72 developed a dynamic scheme based on the configurational bias method<sup>122</sup> that allows to sample between hybridised and free chains on the fly, hence without the need of precomputing  $\Delta G^{\text{cnf}}$ , for every couple of constructs that could potentially bind.

## Acknowledgements

This work was supported by ERC AdG grant 227758 (COLSTRUCTION) and EPSRC Programme Grant EP/I001352/1. S. A-U. acknowledges financial support from the Beijing Municipal Government Innovation Center for Soft Matter Science and Engineering. BMM is supported by the Université Libre de Bruxelles. We acknowledge the present and the past members of the group of Daan Frenkel (P. Varilly, B. Mladeck, M. E. Leunissen, F. J. Martinez-Veracoechea, J. Dobnikar) and of the group of Erika Eiser (L. Di Michele, T. Yanagishima, Z. Ruff) for collaborating with us on the subjects reported by this Perspective.

## References

- 1 C. A. Mirkin, R. C. Letsinger, R. C. Mucic and J. J. Storhoff, A DNA-based method for rationally assembling nanoparticles into macroscopic materials, *Nature*, 1996, **382**, 607.
- 2 A. P. Alivisatos, K. P. Johnsson, X. Peng, T. E. Wilson, C. J. Loweth, M. P. Bruchez and P. G. Schultz, Organization of 'nanocrystal molecules' using DNA, *Nature*, 1996, **382**, 609.
- 3 J. I. Cutler, E. Auyeung and C. A. Mirkin, Spherical nucleic acids, *J. Am. Chem. Soc.*, 2012, **134**(3), 1376–1391.
- 4 S. Y. Park, A. K. R. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz and C. A. Mirkin, DNA-programmable nanoparticle crystallization, *Nat. Mater.*, 2008, **451**, 553–556.
- 5 D. Nykypanchuk, M. M. Maye, D. van der Lelie and O. Gang, DNA-guided crystallization of colloidal nanoparticles, *Nat. Mater.*, 2008, **451**, 549–552.
- 6 M. M. Maye, M. Thilak Kumara, D. Nykypanchuk, W. B. Sherman and O. Gang, Switching binary states of nanoparticle superlattices and dimer clusters by DNA strands, *Nat. Nanotechnol.*, 2010, **5**(2), 116–120.
- 7 Y. Zhang, F. Lu, K. G. Yager, D. van der Lelie and O. Gang, A general strategy for the DNA-mediated self-assembly of



- functional nanoparticles into heterogeneous systems, *Nat. Nanotechnol.*, 2013, **8**(11), 865–872.
- 8 T. Andrew Taton, C. A. Mirkin and R. L. Letsinger, Scano-metric DNA array detection with nanoparticle probes, *Science*, 2000, **289**(5485), 1757–1760.
  - 9 N. L. Rosi, D. A. Giljohann, C. Shad Thaxton, A. K. R. Lytton-Jean, M. S. Han and C. A. Mirkin, Oligonucleotide-modified gold nanoparticles for intracellular gene regulation, *Science*, 2006, **312**(5776), 1027–1030.
  - 10 D. Zheng, D. A. Giljohann, D. L. Chen, M. D. Massich, X.-Q. Wang, H. Iordanov, C. A. Mirkin and A. S. Paller, Topical delivery of siRNA-based spherical nucleic acid nanoparticle conjugates for gene regulation, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**(30), 11975–11980.
  - 11 L. Di Michele and E. Eiser, Developments in understanding and controlling self assembly of DNA-functionalized colloids, *Phys. Chem. Chem. Phys.*, 2013, **15**(9), 3115–3129.
  - 12 N. Geerts and E. Eiser, DNA-functionalized colloids: physical properties and applications, *Soft Matter*, 2010, **6**(19), 4647–4660.
  - 13 P. E. Theodorakis, N. G. Fytas, G. Kahl and C. Dellago, Self-assembly of DNA-functionalized colloids, 2015, arXiv preprint, arXiv:1503.05384.
  - 14 C. Knorowski and A. Travesset, Materials design by DNA programmed self-assembly, *Curr. Opin. Solid State Mater. Sci.*, 2011, **15**(6), 262–270.
  - 15 M. R. Jones, N. C. Seeman and C. A. Mirkin, Programmable materials and the nature of the DNA bond, *Science*, 2015, **347**(6224), 1260901.
  - 16 R. Macfarlane, B. Lee, M. Jones, N. Harris, G. Schatz and C. A. Mirkin, Nanoparticle superlattice engineering with DNA, *Science*, 2011, **334**, 204–208.
  - 17 P. L. Biancianiello, A. J. Kim and J. C. Crocker, Colloidal interactions and self-assembly using DNA hybridization, *Phys. Rev. Lett.*, 2005, **94**, 058302.
  - 18 R. Jin, G. Wu, Z. Li, C. A. Mirkin and G. C. Schatz, What controls the melting properties of DNA-linked gold nanoparticle assemblies?, *J. Am. Chem. Soc.*, 2003, **125**(6), 1643–1654.
  - 19 N. A. Licata and A. V. Tkachenko, Statistical mechanics of DNA-mediated colloidal aggregation, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **74**(4), 041408.
  - 20 A. J. Kim, P. L. Biancianiello and J. C. Crocker, Engineering DNA-mediated colloidal crystallization, *Langmuir*, 2006, **22**(5), 1991–2001.
  - 21 R. Dreyfus, M. E. Leunissen, R. Sha, A. V. Tkachenko, N. C. Seeman, D. J. Pine and P. M. Chaikin, Simple quantitative model for the reversible association of DNA coated colloids, *Phys. Rev. Lett.*, 2009, **102**, 048301.
  - 22 R. Dreyfus, M. E. Leunissen, R. Sha, A. Tkachenko, N. C. Seeman, D. J. Pine and P. M. Chaikin, Aggregation-disaggregation transition of DNA-coated colloids: experiments and theory, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2010, **81**, 041404.
  - 23 P. Varilly, S. Angioletti-Uberti, B. M. Mognetti and D. Frenkel, A general theory of DNA-mediated and other valence-limited colloidal interactions, *J. Chem. Phys.*, 2012, **137**, 094108.
  - 24 B. M. Mognetti, P. Varilly, S. Angioletti-Uberti, F. J. Martinez-Veracoechea, J. Dobnikar, M. E. Leunissen and D. Frenkel, Predicting DNA-mediated colloidal pair interactions, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E378.
  - 25 W. B. Rogers and J. C. Crocker, Direct measurements of DNA-mediated colloidal interactions and their quantitative modeling, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**(38), 15687–15692.
  - 26 M. E. Leunissen and D. Frenkel, Numerical study of DNA-functionalized microparticles and nanoparticles: explicit pair potentials and their implications for phase behavior, *J. Chem. Phys.*, 2011, **134**(8), 084702.
  - 27 S. Angioletti-Uberti, P. Varilly, B. M. Mognetti, A. V. Tkachenko and D. Frenkel, Communication: a simple analytical formula for the free energy of ligand–receptor-mediated interactions, *J. Chem. Phys.*, 2013, **138**, 021102.
  - 28 S. Angioletti-Uberti, P. Varilly, B. M. Mognetti and D. Frenkel, Mobile linkers on DNA-coated colloids: valency without patches, *Phys. Rev. Lett.*, 2014, **113**, 128303.
  - 29 S. Angioletti-Uberti, B. M. Mognetti and D. Frenkel, Reentrant melting as a design principle for DNA-coated colloids, *Nat. Mater.*, 2012, **11**, 518–522.
  - 30 D. Nykypanchuk, M. M. Maye, D. van der Lelie and O. Gang, DNA-based approach for interparticle interaction control, *Langmuir*, 2007, **23**(11), 6305–6314.
  - 31 B. M. Mladek, J. Fornleitner, F. J. Martinez-Veracoechea, A. Dawid and D. Frenkel, Quantitative prediction of the phase diagram of DNA-functionalized nanosized colloids, *Phys. Rev. Lett.*, 2012, **108**(26), 268301.
  - 32 B. M. Mladek, J. Fornleitner, F. J. Martinez-Veracoechea, A. Dawid and D. Frenkel, Procedure to construct a multi-scale coarse-grained model of DNA-coated colloids from experimental data, *Soft Matter*, 2013, **9**, 7342–7355.
  - 33 L. Parolini, B. M. Mognetti, J. Kotar, E. Eiser, P. Cicuta and L. Di Michele, Volume and porosity thermal regulation in lipid mesophases by coupling mobile ligands to soft membranes, *Nat. Commun.*, 2015, **6**, 5948.
  - 34 Y. Zhang, H. Zhou and Z.-C. Ou-Yang, Stretching single-stranded DNA: interplay of electrostatic, base-pairing and base-pair stacking interactions, *Biophys. J.*, 2001, **81**(2), 1133–1143.
  - 35 A more precise definition takes into account the Boltzmann weighted phase space volume of the bound vs. unbound states for the linker. As long as one consider non interacting strands, the two are actually the same.
  - 36 D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Academic Press, 2001, vol. 1.
  - 37 M. S. Wertheim, Fluids with highly directional attractive forces. i. Statistical thermodynamics, *J. Stat. Phys.*, 1984, **35**, 19–34.
  - 38 M. S. Wertheim, Fluids with highly directional attractive forces. ii. Thermodynamic perturbation theory and integral equations, *J. Stat. Phys.*, 1984, **35**, 35–47.



- 39 M. S. Wertheim, Fluids with highly directional attractive forces. iii. Multiple attraction sites, *J. Stat. Phys.*, 1986, **42**, 459–476.
- 40 W. G. Chapman, K. E. Gubbins, C. G. Joslin and C. G. Gray, Theory and simulation of associating liquid mixtures, *Fluid Phase Equilib.*, 1986, **29**, 337–346.
- 41 Y. Wang, Y. Wang, R. Dana, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck and D. J. Pine, Colloids with valence and specific directional bonding, *Nature*, 2012, **491**, 51–55.
- 42 B. Mognetti, M. E. Leunissen and D. Frenkel, Controlling the temperature sensitivity of DNA-mediated colloidal interactions through competing linkages, *Soft Matter*, 2012, **8**, 2213.
- 43 W. B. Rogers and J. C. Crocker, Reply to mognetti *et al.*: DNA handshaking interaction data are well described by mean-field and molecular models, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**(7), E380.
- 44 W. B. Rogers and V. N. Manoharan, Programming colloidal phase transitions with DNA strand displacement, *Science*, 2015, **347**(6222), 639–642.
- 45 M. E. Leunissen, R. Dreyfus, R. Sha, N. C. Seeman and P. M. Chaikin, Quantitative study of the association thermodynamics and kinetics of DNA-coated particles for different functionalization schemes, *J. Am. Chem. Soc.*, 2010, **132**(6), 1903–1913. PMID: 20095643.
- 46 O. Gang, Colloidal self-assembly: melting also on cooling, *Nat. Mater.*, 2012, **11**(6), 487–488.
- 47 U. Rauwald and O. A. Scherman, Supramolecular block copolymers with cucurbit [8] uril in water, *Angew. Chem., Int. Ed.*, 2008, **47**(21), 3950–3953.
- 48 I. Papp, C. Sieben, K. Ludwig, M. Roskamp, C. Böttcher, S. Schlecht A. Herrmann and R. Haag, Inhibition of influenza virus infection by multivalent sialic-acid-functionalized gold nanoparticles, *Small*, 2010, **6**(24), 2900–2906.
- 49 J. Vonnemann, C. Sieben, C. Wolff, K. Ludwig, C. Böttcher, A. Herrmann and R. Haag, Virus inhibition induced by polyvalent nanoparticles of different sizes, *Nanoscale*, 2014, **6**(4), 2353–2360.
- 50 C. Fasting, C. A. Schalley, M. Weber, O. Seitz, S. Hecht, B. Korsch, J. Dornedde, C. Graf, E.-W. Knapp and R. Haag, Multivalency as a chemical organization and action principle, *Angew. Chem., Int. Ed.*, 2012, **51**(42), 10472–10498.
- 51 G. Ercolani, Assessment of cooperativity in self-assembly, *J. Am. Chem. Soc.*, 2003, **125**(51), 16097–16103.
- 52 A. Mulder, J. Huskens and D. N. Reinhoudt, Multivalency in supramolecular chemistry and nanofabrication, *Org. Biomol. Chem.*, 2004, **2**(23), 3409–3424.
- 53 J. Huskens, A. Mulder, T. Auletta, C. A. Nijhuis, M. J. W. Ludden and D. N. Reinhoudt, A model for describing the thermodynamics of multivalent host–guest interactions at interfaces, *J. Am. Chem. Soc.*, 2004, **126**(21), 6784–6797.
- 54 F. J. Martinez-Veracoechea and M. E. Leunissen, The entropic impact of tethering, multivalency and dynamic recruitment in systems with specific binding groups, *Soft Matter*, 2013, **9**(12), 3213–3219.
- 55 Y. Wang, Y. Wang, X. Zheng, T. Ducrot, M.-G. Lee, G.-R. Yi, M. Weck and D. J. Pine, Synthetic strategies toward DNA-coated colloids that crystallize, *J. Am. Chem. Soc.*, 2015, **137**(33), 10760–10766.
- 56 P. S. Randeria, M. R. Jones, K. L. Kohlstedt, R. J. Banga, M. Olvera de la Cruz, G. C. Schatz and C. A. Mirkin, What controls the hybridization thermodynamics of spherical nucleic acids?, *J. Am. Chem. Soc.*, 2015, **137**(10), 3486–3489.
- 57 M. Doyen, K. Bartik and G. Bruylants, DNA-promoted auto-assembly of gold nanoparticles: effect of the DNA sequence on the stability of the assemblies, *Polymers*, 2013, **5**(3), 1041–1055.
- 58 Y. Wang, Y. Wang, X. Zheng, É. Ducrot, J. S. Yodh, M. Weck and D. J. Pine, Crystallization of DNA-coated colloids, *Nat. Commun.*, 2015, **6**, 7253.
- 59 F. W. Starr and F. Sciortino, Model for assembly and gelation of four-armed DNA dendrimers, *J. Phys.: Condens. Matter*, 2006, **18**(26), L347.
- 60 J. Largo, F. W. Starr and F. Sciortino, Self-assembling DNA dendrimers: a numerical study, *Langmuir*, 2007, **23**(11), 5896–5905.
- 61 C. Knorowski and A. Travasset, Dynamics of DNA-programmable nanoparticle crystallization: gelation, nucleation and topological defects, *Soft Matter*, 2012, **8**(48), 12053–12059.
- 62 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, Structural, mechanical and thermodynamic properties of a coarse-grained DNA model, *J. Chem. Phys.*, 2011, **134**(8), 085101.
- 63 B. E. K. Snodin, F. Randisi, M. Mosayebi, P. Sulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, A. A. Louis, *et al.*, Introducing improved structural properties and salt dependence into a coarse-grained model of DNA, 2015, arXiv preprint, arXiv:1504.00821.
- 64 D. M. Hinckley, G. S. Freeman, J. K. Whitmer and J. J. de Pablo, An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: structure, thermodynamics and dynamics of hybridization, *J. Chem. Phys.*, 2013, **139**(14), 144903.
- 65 P. L. Biancaniello, J. C. Crocker, D. A. Hammer and V. T. Milam, DNA-mediated phase behavior of microsphere suspensions, *Langmuir*, 2007, **23**(5), 2688–2693.
- 66 Q. Xu, L. Feng, R. Sha, N. C. Seeman and P. M. Chaikin, Subdiffusion of a sticky particle on a surface, *Phys. Rev. Lett.*, 2011, **106**, 228102.
- 67 R. J. Hunter, *Foundations of colloid science*, 2001.
- 68 F. J. Martinez Veracoechea, B. M. Mognetti, S. Angioletti-Uberti, P. Varilly, D. Frenkel and J. Dobnikar, Designing stimulus-sensitive colloidal walkers, *Soft Matter*, 2014, **10**, 3463–3470.
- 69 J. SantaLucia, A unified view of polymer, dumbbell and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**(4), 1460–1465.
- 70 J. SantaLucia Jr and D. Hicks, The thermodynamics of DNA structural motifs, *Annu. Rev. Biophys. Biomol. Struct.*, 2004, **33**, 415–440.
- 71 D. Bracha, E. Karzbrun, G. Shemer, P. A. Pincus and R. H. Bar-Ziv, Entropy-driven collective interactions in DNA



- brushes on a biochip, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**(12), 4534–4538.
- 72 R. De Gernier, T. Curk, G. V. Dubacheva, R. P. Richter and B. M. Mognetti, A new configurational bias scheme for sampling supramolecular structures, *J. Chem. Phys.*, 2014, **141**(24), 244909.
  - 73 J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, Academic Press, 2006.
  - 74 B. Bozorgui and D. Frenkel, Liquid–vapor transition driven by bond disorder, *Phys. Rev. Lett.*, 2008, **101**(4), 045701.
  - 75 F. J. Martinez-Veracoechea, B. Bozorgui and D. Frenkel, Anomalous phase behavior of liquid–vapor phase transition in binary mixtures of DNA-coated particles, *Soft Matter*, 2010, **6**(24), 6136–6145.
  - 76 F. J. Martinez-Veracoechea, B. M. Mladek, A. V. Tkachenko and D. Frenkel, Design rule for colloidal crystals of DNA-functionalized particles, *Phys. Rev. Lett.*, 2011, **107**, 045902.
  - 77 T. Schmatko, B. Bozorgui, N. Geerts, D. Frenkel, E. Eiser and W. C. K. Poon, A finite-cluster phase in  $\lambda$ -DNA-coated colloids, *Soft Matter*, 2007, **3**(6), 703–706.
  - 78 N. Geerts, T. Schmatko and E. Eiser, Clustering *versus* percolation in the assembly of colloids coated with long DNA, *Langmuir*, 2008, **24**(9), 5118–5123.
  - 79 P. G. Bolhuis, A. A. Louis, J. P. Hansen and E. J. Meijer, Accurate effective pair potentials for polymer solutions, *J. Chem. Phys.*, 2001, **114**(9), 4296–4311.
  - 80 C. Pierleoni, B. Capone and J.-P. Hansen, A soft effective segment representation of semidilute polymer solutions, *J. Chem. Phys.*, 2007, **127**(17), 171102.
  - 81 G. D'Adamo, A. Pelissetto and C. Pierleoni, Coarse-graining strategies in polymer solutions, *Soft Matter*, 2012, **8**(19), 5151–5167.
  - 82 F. Smallenburg and F. Sciortino, Liquids more stable than crystals in particles with limited valence and flexible bonds, *Nat. Phys.*, 2013, **9**(9), 554–558.
  - 83 M. Hadorn, E. Boenzli, K. T. Sørensen, H. Fellermann, P. Eggenberger Hotz and M. M. Hanczyc, Specific and reversible DNA-directed self-assembly of oil-in-water emulsion droplets, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**(50), 20320–20325.
  - 84 L.-L. Pontani, I. Jorjadze, V. Viasnoff and J. Bruijic, Biomimetic emulsions reveal the effect of mechanical forces on cell–cell adhesion, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**(25), 9839–9844.
  - 85 L. Feng, L.-L. Pontani, R. Dreyfus, P. Chaikin and J. Bruijic, Specificity, flexibility and valence of DNA bonds guide emulsion architecture, *Soft Matter*, 2013, **9**(41), 9816–9823.
  - 86 S. A. J. van der Meulen and M. E. Leunissen, Solid colloids with surface–mobile DNA linkers, *J. Am. Chem. Soc.*, 2013, **135**(40), 15129–15134.
  - 87 S. F. Shimobayashi, B. Matteo Mognetti, L. Parolini, D. Orsi, P. Cicuta and L. Di Michele, Direct measurement of DNA-mediated adhesion between lipid bilayers, *Phys. Chem. Chem. Phys.*, 2015, **17**(24), 15615–15628.
  - 88 J. Hu, R. Lipowsky and T. R. Weikl, Binding constants of membrane-anchored receptors and ligands depend strongly on the nanoscale roughness of membranes, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**(38), 15283–15288.
  - 89 G. S. Grest and K. Kremer, Molecular dynamics simulation for polymers in the presence of a heat bath, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1986, **33**(5), 3628.
  - 90 C. Wei Hsu, F. Sciortino and F. W. Starr, Theoretical description of a DNA-linked nanoparticle self-assembly, *Phys. Rev. Lett.*, 2010, **105**(5), 055502.
  - 91 W. Dai, S. K. Kumar and F. W. Starr, Universal two-step crystallization of DNA-functionalized nanoparticles, *Soft Matter*, 2010, **6**(24), 6130–6135.
  - 92 C. Knorowski, S. Burleigh and A. Travesset, Dynamics and statics of DNA-programmable nanoparticle self-assembly and crystallization, *Phys. Rev. Lett.*, 2011, **106**, 215501.
  - 93 C. Knorowski and A. Travesset, Self-assembly and crystallization of hairy (f-star) and DNA-grafted nanocubes, *J. Am. Chem. Soc.*, 2014, **136**(2), 653–659.
  - 94 P. E. Theodorakis, C. Dellago and G. Kahl, A coarse-grained model for DNA-functionalized spherical colloids, revisited: effective pair potential from parallel replica simulations, *J. Chem. Phys.*, 2013, **138**(2), 02510.
  - 95 T. I. Li, R. Sknepnek, R. J. Macfarlane, C. A. Mirkin and M. Olvera de la Cruz, Modeling the crystallization of spherical nucleic acid nanoparticle conjugates with molecular dynamics simulations, *Nano Lett.*, 2012, **12**(5), 2509–2514.
  - 96 T. I. Li, R. Sknepnek and M. Olvera de la Cruz, Thermally active hybridization drives the crystallization of DNA-functionalized nanoparticles, *J. Am. Chem. Soc.*, 2013, **135**(23), 8535–8541.
  - 97 S. Dhakal, K. L. Kohlstedt, G. C. Schatz, C. A. Mirkin and M. Olvera de la Cruz, Growth dynamics for DNA-guided nanoparticle crystallization, *ACS Nano*, 2013, **7**(12), 10948–10959.
  - 98 J. W. Zwanikken, P. Guo, C. A. Mirkin and M. Olvera de la Cruz, Local ionic environment around polyvalent nucleic acid-functionalized nanoparticles, *J. Phys. Chem. C*, 2011, **115**(33), 16368–16373.
  - 99 E. Auyeung, T. I. Li, A. J. Senesi, A. L. Schmucker, B. C. Pals, M. Olvera de la Cruz and C. A. Mirkin, DNA-mediated nanoparticle crystallization into wulff polyhedra, *Nature*, 2014, **505**(7481), 73–77.
  - 100 O. Lee and G. C. Schatz, Molecular dynamics simulation of DNA-functionalized gold nanoparticles, *J. Phys. Chem. C*, 2009, **113**, 2316–2321.
  - 101 O.-S. Lee and G. C. Schatz, Interaction between DNAs on a gold surface, *J. Phys. Chem. C*, 2009, **113**(36), 15941–15947.
  - 102 V. A. Ngo, R. K. Kalia, A. Nakano and P. Vashishta, Super-crystals of DNA-functionalized gold nanoparticles: a million-atom molecular dynamics simulation study, *J. Phys. Chem. C*, 2012, **116**(36), 19579–19585.
  - 103 L. Di Michele, B. M. Mognetti, T. Yanagishima, P. Varilly, Z. Ruff, D. Frenkel and E. Eiser, Effect of inert tails on the thermodynamics of DNA hybridization, *J. Am. Chem. Soc.*, 2014, **136**(18), 6538–6541.





- 104 M. Kenward and K. D. Dorfman, Brownian dynamics simulations of single-stranded DNA hairpins, *J. Chem. Phys.*, 2009, **130**(9), 095101.
- 105 J. P. Lequeieu, D. M. Hinckley and J. J. de Pablo, A molecular view of DNA-conjugated nanoparticle association energies, *Soft Matter*, 2015, **11**, 1919.
- 106 Y. Ding and J. Mittal, Insights into DNA-mediated interparticle interactions from a coarse-grained model, *J. Chem. Phys.*, 2014, **141**(18), 184901.
- 107 It has been reported that dangling sequences can stabilise RNA–RNA duplexes.<sup>123,124</sup> This result may look puzzling in view of the conclusions of ref. 103. It is likely that tails stabilise RNA duplexes by selective interactions as suggested by the fact that, contrarily to ref. 103, the effect is sequence-dependent<sup>123,124</sup>.
- 108 N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye and E. Winfree, On the biophysics and kinetics of toehold-mediated DNA strand displacement, *Nucleic Acids Res.*, 2013, **41**(22), 10641–10658.
- 109 A. A. Louis, Beware of density dependent pair potentials, *J. Phys.: Condens. Matter*, 2002, **14**(40), 9187.
- 110 W. Ostwald, Studien über die bildung und umwandlung fester körper, *Z. Phys. Chem.*, 1897, **22**, 289–330.
- 111 D. Gottwald, G. Kahl and C. N. Likos, Predicting equilibrium structures in freezing processes, *J. Chem. Phys.*, 2005, **122**(20), 204503.
- 112 J. Fornleitner, F. Lo Verso, G. Kahl and C. N. Likos, Genetic algorithms predict formation of exotic ordered configurations for two-component dipolar monolayers, *Soft Matter*, 2008, **4**(3), 480–484.
- 113 A. Travesset, Binary nanoparticle superlattices of soft-particle systems, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**(31), 9563–9567.
- 114 F. Varrato, L. Di Michele, M. Belushkin, N. Dorsaz, S. H. Nathan, E. Eiser and G. Foffi, Arrested demixing opens route to bigels, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**(47), 19155–19160.
- 115 L. Di Michele, F. Varrato, J. Kotar, S. H. Nathan, G. Foffi and E. Eiser, Multistep kinetic self-assembly of DNA-coated colloids, *Nat. Commun.*, 2013, **4**, 2007.
- 116 C. B. Carlson, P. Mowery, R. M. Owen, E. C. Dykhuizen and L. L. Kiessling, Selective tumor cell targeting using low-affinity, multivalent interactions, *ACS Chem. Biol.*, 2007, **2**(2), 119–127.
- 117 C. W. Cairo, J. E. Gestwicki, M. Kanai and L. L. Kiessling, Control of multivalent interactions by binding epitope density, *J. Am. Chem. Soc.*, 2002, **124**(8), 1615–1619.
- 118 F. J. Martinez-Veracoechea and D. Frenkel, Designing super selectivity in multivalent nano-particle binding, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**(27), 10963–10968.
- 119 L. Ronald and G. Treloar, The statistical length of long-chain molecules, *Rubber Chem. Technol.*, 1946, **19**(4), 1002–1008.
- 120 C. D. Wick and J. I. Siepmann, Self-adapting fixed-end-point configurational-bias monte carlo method for the regrowth of interior segments of chain molecules with strong intramolecular interactions, *Macromolecules*, 2000, **33**(19), 7207–7218.
- 121 G. C. A. M. Mooij and D. Frenkel, The overlapping distribution method to compute chemical potentials of chain molecules, *J. Phys.: Condens. Matter*, 1994, **6**(21), 3879.
- 122 J. I. Siepmann and D. Frenkel, Configurational bias monte carlo: a new sampling scheme for flexible chains, *Mol. Phys.*, 1992, **75**(1), 59–70.
- 123 T. Ohmichi, S.-I. Nakano, D. Miyoshi and N. Sugimoto, Long RNA dangling end has large energetic contribution to duplex stability, *J. Am. Chem. Soc.*, 2002, **124**(35), 10367–10372.
- 124 S. Limmer, H.-P. Hofmann, G. Ott and M. Sprinzl, The 3'-terminal end (ncca) of tRNA determines the structure and stability of the aminoacyl acceptor stem, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**(13), 6199–6202.

