Theoretical Description of a DNA-Linked Nanoparticle Self-Assembly

Chia Wei Hsu,[1] Francesco Sciortino,[2] and Francis W. Starr[1]

[1]*Department of Physics, Wesleyan University, Middletown, Connecticut 06459, USA*

[2]*Dipartimento di Fisica and CNR-ISC, Università di Roma La Sapienza, Piazzale Aldo Moro 2, I-00185 Rome, Italy*

(Received 3 May 2010; published 30 July 2010)

Nanoparticles tethered with DNA strands are promising building blocks for bottom-up nanotechnology, and a theoretical understanding is important for future development. Here we build on approaches developed in polymer physics to provide theoretical descriptions for the equilibrium clustering and dynamics, as well as the self-assembly kinetics of DNA-linked nanoparticles. Striking agreement is observed between the theory and molecular modeling of DNA-tethered nanoparticles.

DOI: 10.1103/PhysRevLett.105.055502 PACS numbers: 81.05.Zx, 81.16.Dn, 81.16.Fg

The specificity, directionality, and technological control over base sequence make DNA an attractive linking unit for artificial constructs [1–3]. One such approach is to attach DNA strands of designed base sequence onto a nanoparticle (NP), thereby creating ''molecules'' that self-assemble into highly organized structures through complementary pairings of DNA [4,5]. Recent experiments have demonstrated that uniformly DNA-functionalized NPs can self-assemble into disordered [6–9] or ordered crystal structures [10–13]. NPs with a discrete number of attachments are harder to prepare, but they provide possibilities for even more diverse structures [14–17]. In addition to examining the possible equilibrium structures, it is also vital to develop an understanding for the kinetics of the self-assembly process [8,12,13,18,19], since the pathway to desired structures may depend sensitively on sample preparation.

In this Letter, we build on concepts from polymer physics to provide a theoretical framework that describes both the equilibrium properties and the self-assembly kinetics of NPs functionalized with a small number of ssDNA. We demonstrate the applicability of this theory to molecular simulations of a binary mixture where NPs are functionalized with two or three ssDNA, similar to experimentally realized systems [15].

The coarse-grained molecular model we use for the DNA-functionalized NP is a modest modification of a previous model [20]. The inset in Fig. 1 provides a graphical representation. The sugar-phosphate backbone is represented by connected beads with only excluded volume interactions; each bead carries an additional ''sticky'' site that represents a base (A, T, C, or G) and can bond only with another sticky site of complementary type (A-T or C-G pair). The size of the sticky site is small relative to the backbone beads, so that a base can bond to at most one other complementary base. A bending potential between consecutive triplets of backbone beads is included to model the characteristic rigidity of the DNA strands and to maintain the angle between ssDNA attached to the same core (120° for threefold-functional NPs and 180° for twofold-functional NPs). A more complete description of the model is provided in Ref. [20]. We consider ssDNA four bases in length with the sequence A-C-G-T, chosen to enable complementary pairing between ssDNA on different NPs. To mimic solvent effects on dynamics, we simulate this model by using dissipative particle dynamics [21]. Length is measured in diameter of the beads, and temperature T is measured in units of the bonding energy between complementary bases, taking Boltzmann's constant $k_B = 1$ for simplicity.

We consider a system with $N_2 = 945$ twofold-functional particles and $N_3 = 55$ threefold-functional particles. The presence of threefold-functionalized particles facilitates branching of otherwise linear chains, thereby opening the possibility to form a percolating network and phase separation at low density [22]. For part of our study where system size is less important, we consider a smaller system with $N_2 = 189$ and $N_3 = 11$ so that we may more readily study the dynamics of the most slowly relaxing systems. Nonetheless, equilibration at the lowest T studied still requires 5×10^9 integration steps—roughly 3 months of

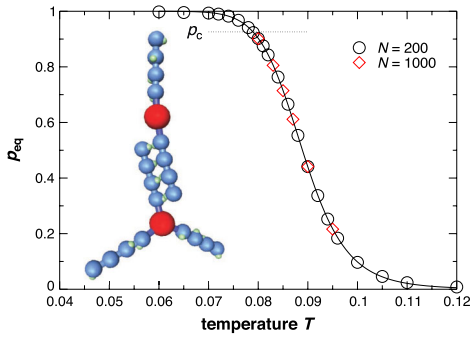

FIG. 1 (color online). Temperature dependence of the fraction of bonded arms at equilibrium, p_{eq} for system sizes $N = 200$ and $N = 1000$. The solid line represents Eq. (1) with $\Delta U = 1.78$ and $\Delta S = 20.0$. The dotted line indicates percolation threshold predicted by FS theory. The inset shows a graphical representation of one threefold-functional particle bonding to one twofold-functional particle. Large red spheres indicate NPs, blue spheres indicate backbone beads of ssDNA, and small green spheres indicate sticky sites.

computation on a single core with current resources. The number density is fixed at $(N_2 + N_3)/V = 0.0065$ (corresponding to mean separation of 5.4 bases between particles); we have confirmed (using the structure factor) that there is no (gas-liquid) phase separation at this density.

In this system, the extent of reaction p is characterized by the fraction of bonded strands. We define two strands as bonded when half or more of the base pairs on them are linked (i.e., the base-base potential energy is negative). The T dependence of equilibrium p is well described (see Fig. 1 and Ref. [20]) by a two-state expression

$$p_{\mathrm{eq}}(T) = \left[1 + \exp\left(-\frac{\Delta U - T\Delta S}{k_B T}\right)\right]^{-1}, \qquad (1)$$

where ΔU and ΔS measure, respectively, the energy and entropy change associated with the formation of a double strand.

To quantify the equilibrium assembly, we first calculate the number of clusters N_k of size k at each T [Fig. 2(a)]. To describe our results, we consider the theoretical predictions of Flory [23] and Stockmayer [24] (FS). The FS theory assumes that no intracluster loops are present and that every functional group is equally reactive. The theory predicts the most probable number of clusters N_{nl} containing n threefold-functional particles and l twofold-functional particles is

$$N_{nl}(p) = N_3(1-p)^{n+2}(p_3 p)^{n-1}(p_2 p)^l w_{nl}, \qquad w_{nl} = 3\frac{(2n+l)!}{n!l!(n+2)!}, \qquad (2)$$

where $p_3 = 3N_3/(2N_2 + 3N_3)$ and $p_2 = 1 - p_3$ are the portions of functional groups belonging to threefold-functional or twofold-functional particles, respectively. A sum yields the analytic prediction $N_k = \sum_{n+l=k} N_{nl}$. Since the only variable in Eq. (2) is the extent of reaction p (which we already know), we have a parameter-free description of N_k. We find quantitative agreement between this prediction and our results [Fig. 2(a)]. We can also evaluate the number of finite-size clusters $N_c = \sum_k N_k$, which we compare with the corresponding simulation results in Fig. 2(b). We again find quantitative agreement in the whole range of p—both below and above the percolation threshold $p_c = 1/(1 + p_3) = 0.9256$—a demonstration that in this system an accurate analytic description for the equilibrium clustering is possible. Qualitatively, these clusters are dominated by long chains of twofold-functional NP, with occasional links provided by the threefold-functional NP.

The clustering process has a profound effect on the dynamics of the system, as revealed by the diffusion constant D, which drops several orders of magnitude in the small T region where clusters form. Figure 3(a) shows D evaluated from the asymptotic behavior of the mean-square displacement $\langle \Delta r^2 \rangle = 6Dt$ as a function of the inverse temperature $1/T$. Close to the percolation transition, D takes on an Arrhenius T dependence.

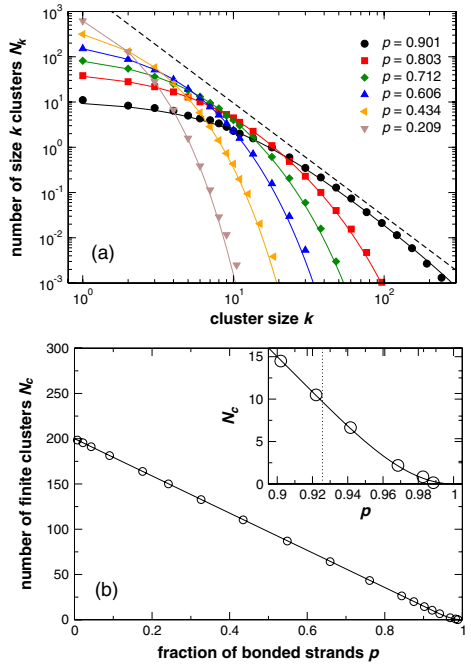

FIG. 2 (color online). (a) Equilibrium distribution of finite-size clusters N_k. Points are simulation data (system size $N = 1000$), and lines are the FS predictions. The dashed line represents the expected power-law decay $N_k \sim k^{-2.5}$ at percolation [31]. (b) Number of finite-size clusters N_c as a function of p. Points are simulation data (system size $N = 200$), and the black solid line is the FS prediction. The inset enlarges the region above percolation threshold (indicated by the dotted line).

To understand the behavior of D, consider that the overall diffusion constant is an average over particles belonging to clusters of different sizes; therefore $D = \sum_k D_k(kN_k)/N$, where D_k is the diffusion constant for particles in a cluster of size k [25]. Typically, D_k drops inversely to the cluster size, so that $D_k \approx D_1/k$, the so-called Stokesian limit of diffusion [26,27]. In this approximation, $D \approx D_1 \sum_k N_k/N = D_1 N_c/N$. If we take the monomer diffusion coefficient D_1 as a constant, we can readily test this prediction. Figure 3(b) shows that D obeys this prediction well, with small deviations at the largest and smallest values. Numerically, we find $D_1 = 0.145$, roughly equal to the high-T asymptotic value of D, where monomers dominate. As shown in Fig. 2(b), $N_c \sim (1 - p)$ for $p < p_c$, and so we also expect $D \sim (1 - p)$ to leading order [Fig. 3(a), inset]. Since at low T, $(1 - p) \sim \exp(-\Delta U/k_B T)$ [from Eq. (1)], we can explain the Arrhenius behavior $D \sim \exp(-\Delta U/k_B T)$ in the low-T region. Indeed, ΔU obtained from $p(T)$ (Fig. 1) matches that for D, demonstrating consistency.

Having successfully described the clustering and dynamics at equilibrium, we next explore to what degree the self-assembly kinetics can be analytically predicted. We focus on the evolution of the system after a temperature jump (at $t = 0$) from a high-T unassociated state to a low-T self-assembled state, following the evolution of p from its

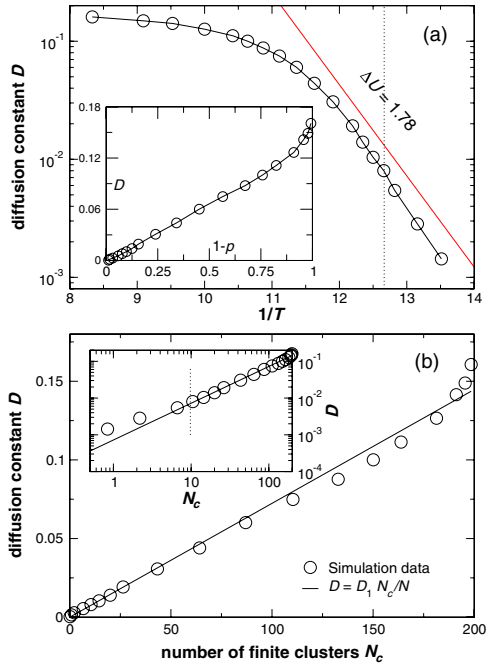

FIG. 3 (color online). (a) Diffusion constant D as a function of inverse temperature. The dotted line indicates the location of the percolation threshold, and the red solid line represents Arrhenius behavior with an activation energy $\Delta U = 1.78$—the same value reported for $p(T)$ in Fig. 1. The inset shows D as a function of fraction of nonbonded arms. (b) Diffusion constant D versus number of finite clusters N_c. The line shows the predicted behavior $D = D_1 N_c/N$. The calculation of $\langle \Delta r^2 \rangle$ is averaged over 8–15 independent runs.

initial value $p(t = 0) \approx 0$ to its final equilibrium value $p(t = \infty) = p_{\mathrm{eq}}(T)$. With the assumption that every strand is equally reactive independent of the size of cluster to which it attaches, $p(t)$ satisfies [28]

$$\frac{dp}{dt} = p_{\mathrm{eq}} k_{\mathrm{overall}} \left[\frac{(1-p)^2}{(1-p_{\mathrm{eq}})^2} - \frac{p}{p_{\mathrm{eq}}} \right], \tag{3}$$

where k_{overall} is the overall rate coefficient for bond breaking. Equation (3) simply states that the rate of bond formation (first term) depends only on the probability of finding two unassociated strands and that the rate of bond fragmentation (second term) depends only on the probability of finding a bonded strand. When this kinetic process is dominated by the intrinsic rate of reaction (i.e., dsDNA formation), $k_{\mathrm{overall}} = k_c$ is a constant, and Eq. (3) can be integrated analytically to give the time evolution

$$p(t) = p_{\mathrm{eq}} \frac{1 - \Lambda e^{-\Gamma t}}{1 - p_{\mathrm{eq}}^2 \Lambda e^{-\Gamma t}}, \tag{4}$$

where $\Gamma = k_c(1 + p_{\mathrm{eq}})/(1 - p_{\mathrm{eq}})$ determines the rate of reaction, and $\Lambda = [1 - p(0)/p_{\mathrm{eq}}]/[1 - p(0)p_{\mathrm{eq}}]$ incorporates the initial condition $p(0)$. It has been shown by van Dongen and Ernst [28] that, in this so-called "reaction-limited" or "chemical-limited" process, Eq. (3) provides a time-dependent distribution $N_k(p(t))$ [together with Eq. (2)] that satisfies the Smoluchowski rate equation with condensation and fragmentation terms below percolation [29].

Using our simulation results, we test if our system can be described by the reaction-limited description given by Eq. (3) with $k_{\mathrm{overall}} = k_c$. In Fig. 4, we see that the reaction-limited expression describes the rate of bond dp/dt and time evolution $p(t)$ well for most values of p but fails when $p \lesssim 0.3$. This indicates that there is a rate-controlling factor for small p not captured by a simple reaction-limited description. The trend in dp/dt shows that this additional rate-controlling factor slows down the overall reaction rate at small p. To incorporate the effect, we introduce an additional reaction rate $k_{\mathrm{small}} = k_0 p$ that changes linearly with p. Since reaction rates sum inversely [26], $k_{\mathrm{overall}}^{-1} = k_c^{-1} + (k_0 p)^{-1}$. Figure 4 shows that this modified k_{overall} accurately describes our simulation results for the assembly kinetics over the entire self-assembly process. By numerically integrating Eq. (3) using the appropriate initial condition $p(0)$, we also provide an accurate estimate of the time evolution $p(t)$ based on this expression. Surprisingly, this prediction works well even for the case in which the final state percolates ($T = 0.06$). The inset in Fig. 4 shows that both rate constants k_c and k_0 follow an Arrhenius behavior (i.e., $\ln k \sim 1/T$).

The presence of a second rate constant that is dominant at small p may be a consequence of the fact that bonding NP via dsDNA requires two steps: (i) the strands of opposing NPs must be properly oriented, and (ii) once oriented, the linking of individual bases must occur. Since base-pair bonds are dynamic, there is a constant "flickering" of bonds. Thus, when p is significant (here, apparently $p \gtrsim 0.3$), the flickering should be a consequence of the regular breaking and reforming of individual base-pair bonds. As a result, the strands will typically remain oriented to facilitate base-pair bonding, making the rate constant associated with orientation irrelevant (i.e., k_{overall} dominated by k_c). For very small p, the intermittency of the bonds is large enough that strands can "wander," and as a result the time needed to orient two strands becomes the rate limiting factor, and so k_{small} dominates. This is consistent with the fact that for simpler systems, where arm orientation is not a factor, there is only a single rate constant to consider [30]. The effect of this mechanism is the opposite of the diffusion-limited scenario, which is usually important only for long-time dynamics, when p is large [26,29].

Our studies show that, by combining a thermodynamic description for $p_{\mathrm{eq}}(T)$, the FS theory for $N_k(p)$, an approximate expression for D, and the rate equation with appropriate rate constants, we can provide a comprehensive theoretical description for the clustering, equilibrium dynamics, and assembly kinetics that accurately predicts the behavior of this DNA-functionalized NP system. The complex process of self-assembly through dsDNA formation

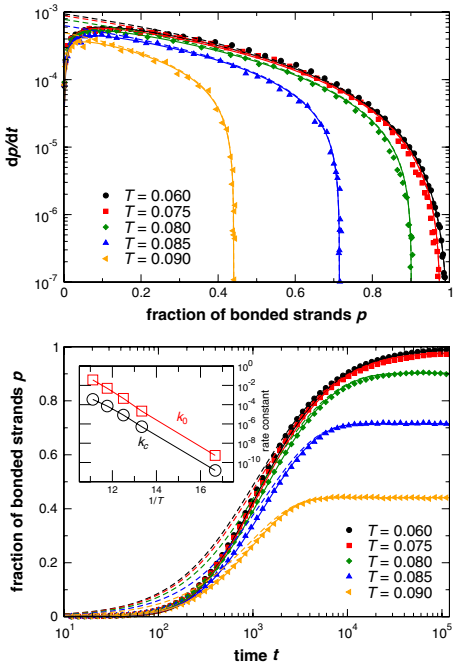

FIG. 4 (color online). (a) Rate of bond formation dp/dt as a function of fraction p of bonded strands, and (b) time evolution of p starting from a high-T state [$T_{\text{init}} = 0.2$, $p(0) = 0.008$]. The inset shows the T dependence of the two rate constants. Symbols are simulation results (system size $N = 1000$), averaged over 35 independent runs at short time and 15 runs at longer time. Dashed and solid lines represent, respectively, the prediction of Eq. (3) in the reaction-limit scenario (constant k) and with consideration of the pk_0 rate coefficient.

can be simplified as a chemical-limited reaction plus an additional reaction rate $k_0 p$ associated with strand orientation. An obvious question is whether the results obtained for our simplified DNA simulation model will be transferable to real systems. The experimentally synthesized systems studied by Alivisatos and co-workers [15] offer the possibility to directly test our predictions. Additionally, our approach could be applicable to the experiments of Gang and co-workers [8] where there are many strands attached to a core NP, but only a small number of which are available for binding. They also consider micron-sized colloidal cores in that work, which can potentially significantly alter the assembly dynamics.

The successful theoretical description of our numerical results shows that existing frameworks can be adapted to describe novel self-assembled systems. The next step in this development is the satisfactory inclusion of closed-loop structures, which is important for applications when the number of functionalizing strands is larger.

We thank J. Douglas for discussion and Wesleyan University for computer time, which was supported by National Science Foundation Grant No. CNS-0959856. This work was supported by National Science Foundation Grant No. DMR-0427239. F. S. acknowledges support from ERC-226207-PATCHYCOLLOIDS and NoE SoftComp NMP3-CT-2004-502235.

[1] C. M. Niemeyer, Curr. Opin. Chem. Biol. **4**, 609 (2000).
[2] N. C. Seeman, Nature (London) **421**, 427 (2003).
[3] A. Condon, Nat. Rev. Genet. **7**, 565 (2006).
[4] C. A. Mirkin, R. L. Letsinger, R. C. Mucic, and J. J. Storhoff, Nature (London) **382**, 607 (1996).
[5] A. P. Alivisatos *et al.*, Nature (London) **382**, 609 (1996).
[6] V. T. Milam *et al.*, Langmuir **19**, 10 317 (2003).
[7] N. C. Harris and C.-H. Kiang, Phys. Rev. Lett. **95**, 046101 (2005).
[8] M. M. Maye, D. Nykypanchuk, D. van der Lelie, and O. Gang, Small **3**, 1678 (2007).
[9] N. Geerts, T. Schmatko, and E. Eiser, Langmuir **24**, 5118 (2008); N. Geerts and E. Eiser, Soft Matter **6**, 664 (2010).
[10] S. Y. Park *et al.*, Nature (London) **451**, 553 (2008).
[11] D. Nykypanchuk, M. M. Maye, D. van der Lelie, and O. Gang, Nature (London) **451**, 549 (2008); H. Xiong, D. van der Lelie, and O. Gang, Phys. Rev. Lett. **102**, 015504 (2009).
[12] A. J. Kim *et al.*, Nature Mater. **8**, 52 (2009).
[13] R. J. Macfarlane *et al.*, Proc. Natl. Acad. Sci. U.S.A. **106**, 10 493 (2009).
[14] K. M. Stewart and L. W. McLaughlin, J. Am. Chem. Soc. **126**, 2050 (2004); K. M. Stewart, J. Rojo, and L. W. McLaughlin, Angew. Chem. **43**, 5808 (2004).
[15] D. Zanchet *et al.*, Nano Lett. **1**, 32 (2001); S. A. Claridge *et al.*, Nano Lett. **8**, 1202 (2008).
[16] K. Suzuki, K. Hosokawa, and M. Maeda, J. Am. Chem. Soc. **131**, 7518 (2009).
[17] C. W. Hsu, J. Largo, F. Sciortino, and F. W. Starr, Proc. Natl. Acad. Sci. U.S.A. **105**, 13 711 (2008); W. Dai, C. W. Hsu, F. Sciortino, and F. W. Starr, Langmuir **26**, 3601 (2010).
[18] P. L. Biancaniello, A. J. Kim, and J. C. Crocker, Phys. Rev. Lett. **94**, 058302 (2005); Biophys. J. **94**, 891 (2008).
[19] R. Dreyfus *et al.*, Phys. Rev. Lett. **102**, 048301 (2009).
[20] F. W. Starr and F. Sciortino, J. Phys. Condens. Matter **18**, L347 (2006); J. Largo, F. W. Starr, and F. Sciortino, Langmuir **23**, 5896 (2007).
[21] R. D. Groot and P. B. Warren, J. Chem. Phys. **107**, 4423 (1997).
[22] E. Bianchi, P. Tartaglia, E. La Nave, and F. Sciortino, J. Phys. Chem. B **111**, 11 765 (2007).
[23] P. Flory, *Principles of Polymer Chemistry* (Cornell University Press, New York, 1953).
[24] W. H. Stockmayer, J. Chem. Phys. **11**, 45 (1943).
[25] A. Fierro, T. Abete, and A. Coniglio, J. Chem. Phys. **131**, 194906 (2009).
[26] G. Oshanin and M. Moreau, J. Chem. Phys. **102**, 2977 (1995).
[27] J. D. Guzman, R. Pollard, and J. D. Schieber, Macromolecules **38**, 188 (2005).
[28] P. G. J. van Dongen and M. H. Ernst, J. Stat. Phys. **37**, 301 (1984).
[29] S. Corezzi *et al.*, J. Phys. Chem. B **114**, 3769 (2010).
[30] F. Sciortino *et al.*, Soft Matter **5**, 2571 (2009).
[31] M. Rubinstein and R. H. Colby, *Polymer Physics* (Oxford University Press, New York, 2003).