Check for updates

Hyperbranched DNA clusters†

Enrico Lattuada, [ID] * Debora Caprara, [ID] Vincenzo Lamberti [ID] and Francesco Sciortino [ID]

Taking advantage of the base-pairing specificity and tunability of DNA interactions, we investigate the spontaneous formation of hyperbranched clusters starting from purposely designed DNA tetravalent nanostar monomers, encoding in their four sticky ends the desired binding rules. Specifically, we combine molecular dynamics simulations and Dynamic Light Scattering experiments to follow the aggregation process of DNA nanostars at different concentrations and temperatures. At odds with the Flory–Stockmayer predictions, we find that, even when all possible bonds are formed, the system does not reach percolation due to the presence of intracluster bonds. We present an extension of the Flory–Stockmayer theory that properly describes the numerical and experimental results.

In recent years, the relationship between reversible self-assembly of patchy colloidal particles and irreversible aggregation of chemical units is receiving a renewed interest.[1,2] This connection has been nourished by the observation that the clusters considered in the Wertheim theory for associating liquids[3,4] are the same loopless clusters considered in the Flory–Stockmayer (FS) theory of polyfunctional condensation.[5,6] For example, the conditions for the formation of infinite networks, valid for describing the self-condensation of f-functional A_f monomers (*i.e.* monomers with a number of f distinct reactive groups), are equivalent to the ones for colloidal particles with f attractive patches. In contrast to the chemical case, for which equilibrium conditions are assumed – but hardly realizable in experiments due to the covalent nature of the bonds – colloidal aggregation may proceed to equilibrium, allowing for more accurate control of the theoretical FS predictions.

For the case of colloidal particles with f patches, there is evidence that the range of validity of the FS predictions becomes wider and wider upon decreasing f.[7,8] For binary mixtures of very small average "valence" $\langle f \rangle$, the FS predictions provide a quite accurate description of the cluster size distribution, except for bond probabilities very close to the percolation threshold (which is properly predicted theoretically).[1] For this reason, patchy colloidal particles have become a test ground for revisiting old FS predictions under proper equilibrium conditions. Furthermore, they provided a way to access the role of the bonding loops (*i.e.* close paths of bonds), which are commonly neglected both in FS and Wertheim theories.

An interesting aggregation phenomenon takes place in one-component systems made of AB_{f-1} monomers‡ (*i.e.* where each monomer is composed of one reactive group of type A and $f - 1$ of type B), where A condenses with B, but reactions between identical functional groups (AA and BB) are forbidden. The clusters resulting from this aggregation process are commonly known as hyperbranched polymers, a term introduced by Kim and Webster in their works on the synthesis of highly branched polymers.[9,10] In the last decades, the interest towards the synthesis and understanding of these materials has continuously grown, representing a challenge for innovative applications. Hyperbranched polymers constitute an appealing alternative to dendrimers, owing to their facile synthesis and high tunability.[11,12] Similar to other branched polymers, they are characterized by high exposure of functional groups, three-dimensional globular structure, low viscosity, and good solubility.[13,14] Potential applications include surface coating,[15] composite material filling to increase the thermal and mechanical stability,[16] drug and gene delivery,[17–21] nanoparticle grafting for diagnostic imaging to reduce the toxicity,[22,23] and their use in sensors.[24–26]

From the theoretical standpoint, the aggregation of AB_{f-1} units is particularly interesting for several reasons: (i) it is analytically tractable (neglecting the formation of closed bond loops); (ii) the cluster size distributions for branched polymers are requested as an intermediate step in the evaluation of several polyfunctional condensation processes; and (iii) it gives rise, according to the FS theory, to an aggregation phenomenon in which the fully bonded case (when all A groups have reacted) corresponds to the percolation transition. Therefore,

Physics Department, Sapienza University, P.le Aldo Moro 5, 00185 Rome, Italy.
E-mail: enrico.lattuada@uniroma1.it

† Electronic supplementary information (ESI) available. See DOI: 10.1039/D0NR04840B

‡ The AB_{f-1} monomer can be also indicated as A−R−B_{f-1} or ARB_{f-1}.

hyperbranched polymers do not have a gel phase but only a sol one.

Despite the interest towards hyperbranched polymers, an accurate comparison of the FS theoretical predictions with numerical and experimental results has rarely been attempted.[27,28] Actually, the hypothesis of the absence of intramolecular reactions is expected to get progressively worse for large degrees of polymerization. In fact, an unreacted A will most likely interact with one of the (nearby) B sites belonging to the same cluster, hence forming a closed loop. As a result, the cluster size distribution of hyperbranched polymers may not follow the FS predictions.

In line with the conceptual correspondence between patchy colloids and reactive monomers, we present here a combined numerical and experimental study of a colloidal analog of the AB_{f-1} hyperbranching condensation. In particular, we design specific DNA oligomers able to self-assemble into bulk quantities of identical four-armed (i.e. $f = 4$) AB_3 particles which can interact in a controlled way.[29–32] We exploit molecular dynamics (MD) simulations based on the oxDNA2 coarse-grained interaction potential[33,34] to follow the particle aggregation process and to compare the numerical results with theoretical predictions. This numerical study allows us to estimate the role of intramolecular binding and how the presence of closed loops modifies the cluster size distribution. We then demonstrate that a cluster-based thermodynamic treatment, which also includes intracluster bonds, can be developed to extend the FS theory in order to accurately describe the simulation data and to provide a significant guide to the experimental results. Finally, we realize the same system in the laboratory and follow the aggregation process via Dynamic Light Scattering (DLS) to provide evidence that the presence of intracluster bonds prevents the formation of a percolating state, even in the limit of full bonding.

The DNA particle

In the last years, the ideas that have emerged with regard to DNA nanotechnology[35,36] have been borrowed by the soft matter community to synthesize bulk quantities of nanometric particles with controlled shape and binding properties,[29,30,32,37,38] including dendrimers.[12] The particles we envision in this study are assembled starting from four distinct single-stranded DNA oligomers, each composed of 52 nucleotides, containing properly designed sequences of complementary groups (see Fig. 1a). The self-assembly of these strands generates a well-defined nanostar (NS), composed of four double-stranded arms of 20 base pairs (Fig. 1b and d). Two unpaired adenines, located at the center of each of the four single-stranded sequences composing the NS (resulting in a total of eight unpaired bases, which form the NS core) provide arm flexibility. Each arm terminates with a 9-base long single-stranded sticky sequence preceded by an additional unbonded adenine, which is inserted to ease the linking between different NSs.

Most of the previous works on DNA NSs focused on the A_4 case, tetra-armed NSs with identical self-complementary sticky
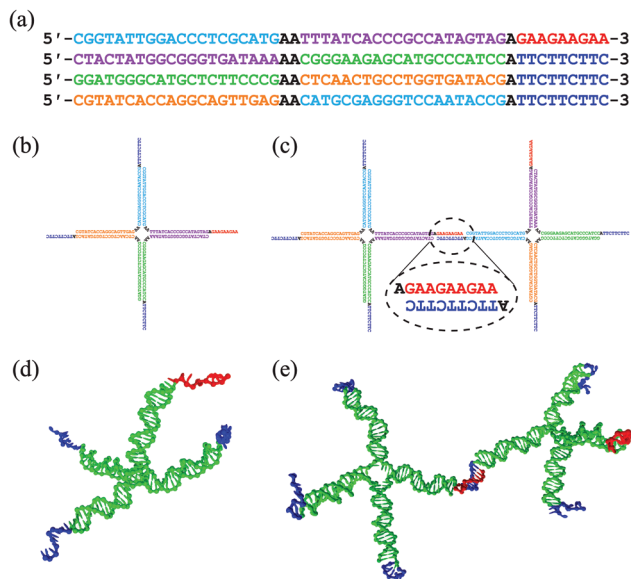
Fig. 1 (a) Oligonucleotide composition of the four strands comprising the tetravalent monomer. Colors correspond to sequence pairings forming the double-stranded monomer arms, shown in (b). The last nine bases are the A (red) and B (blue) sticky-end sequences, respectively. (c) NS–NS binding via the 9-base long DNA sequences, located at the tips. (d and e) Corresponding images generated from oxDNA configurations.

sequences, originating all identical AA bonds.[29–31,39–47] The A_4 system shows the analog of the gas–liquid phase separation at low temperatures, driven by the association of the self-complementary sticky ends. For concentrations larger than the "liquid" coexistence value, the system forms a reversible equilibrium gel, which exhibits the static and dynamic features expected for colloidal particles of valence four.[29,30]

With the specific design presented here, an individual NS possesses one sticky end of type A and three of type B in order to mimic a tetravalent AB_3 monomer. The A and B sticky regions have been adequately selected to allow only AB bonds between different NSs (Fig. 1c and e).

The temperature response of the system is strictly related to the number of nucleotides in the complementary sequences, whose length allows to distinguish different hierarchical self-assembly processes. As shown in Fig. S1 of the ESI,† above T_{NS} ≈ 77 °C; the sample is composed of single and freely diffusing DNA strands. Around T_{NS}, the complementary sequences comprising the double-stranded arms start to self-assemble, giving rise to a solution of unbonded NSs. On further cooling, around T_b ≈ 42 °C, the sticky ends start to pair forming inter-, and possibly intra-, NS AB bonds. At ambient temperature and below, essentially all possible AB bonds are formed.

Materials and methods

Numerical methods

To simulate the aggregation kinetics of AB_3 DNA NSs, we employ the coarse-grained model oxDNA2, which is able to

reproduce the structural and thermodynamic properties of single- and double-stranded DNA molecules.[33,34] The interactions between nucleotides account for the excluded volume, backbone connectivity, Watson–Crick hydrogen bonding, stacking, cross-stacking, coaxial-stacking, and for electrostatic interactions at salt concentrations $c_{Na^+} > 0.1$ M. The model parameters have been adjusted to reproduce the experimental melting temperature data.[33,48,49] A code implementing the oxDNA2 model is freely available.[50]

Initial configurations are generated by randomly placing copies of an already assembled DNA tetramer in the simulation box, provided there is no overlap between the nucleotides (for more details, see section B of the ESI†). The volume is computed for the different values of N and concentrations (c = 2 mg ml^{-1}, 10 mg ml^{-1}, and 20 mg ml^{-1}, equal to those experimentally investigated), using a tetramer molecular weight of M_w = 63 893 g mol^{-1}. We perform MD simulations in the NVT ensemble with systems consisting of N = 300, 1000, and 2000 DNA NSs of AB$_3$ type. The largest system corresponds to ~4 × 10^5 interaction sites. The temperature in the simulation, kept at T = 45 °C, is enforced using an Anderson-like thermostat that emulates a Brownian motion.[51] For the selected temperature, we are able to follow the equilibration of the system for up to 10^{10} MD time steps (corresponding to ~30 μs of real time and six months of continuous computation per run), taking advantage of the computational power of modern Graphic Processing Units (GPU).

Experimental methods

DNA sequences are purchased from Integrated DNA Technologies (IDT) with PAGE purification. Lyophilized samples are initially dispersed in a filtered, DNAse-free, 50 mM NaCl solution. Tetravalent particles are pre-assembled by mixing equimolar quantities of the single-stranded components to a final NS concentration c = 22 mg ml^{-1} (344 μM). The mixture is heated up to 90 °C, incubated for 20 minutes, and slowly cooled down to room temperature overnight. NS annealing is carried out using a Memmert oven.

We experimentally investigate three samples at different NS concentrations: 2 mg ml^{-1}, 10 mg ml^{-1}, and 20 mg ml^{-1}. The samples are prepared in borosilicate glass capillaries (inner diameter 2.4 mm, Hilgenberg GmbH). We dilute the NS batch suspension with a NaCl solution at a proper molarity, in order to obtain 30 μl of each sample at a final NaCl concentration of 250 mM. Finally, we cover the suspension with 20 μl of silicone oil and seal the capillary using UV resin to avoid sample evaporation.

DLS measurements are carried out at a fixed angle θ = 90° with a custom-made setup consisting of a 633 nm He–Ne Laser (17 mW, Newport Corp.) and a multi-tau digital correlator (Brookhaven Inst.) connected to a multi-mode optical fiber. The scattered light impinging on the fiber is spatially filtered, resulting in a high coherence factor ($\beta \simeq 0.8$). Samples are immersed in a water bath connected to a thermostat. The actual temperature of the bath near the sample is measured using a thermocouple probe with a ΔT = ±0.05 °C accuracy.

For each selected temperature, within the interval 10 °C ≤ T ≤ 50 °C every $\Delta T \simeq$ 5 °C, the sample is thermalized for 40 minutes before starting the acquisition. Each measurement lasts 10 minutes. The autocorrelation functions of the scattered intensity $g_2(t)$ are calculated from the correlator output and converted into the field correlation functions $g_1(t)$ using the Siegert relation.[52]

Results

Numerical results

We start by examining the number of bonds $\#_b(t)$ as a function of time. We assume two NSs to be bound when at least five complementary bases of the AB sticky sequences hybridize (see section C of the ESI† for more details). Since the maximum number of possible bonds (fully bonded state) is equal to the number of particles N, the fraction of bonds $\#_b(t)/N$ coincides with the probability $p_b(t)$ that an arbitrary A sticky end is engaged in a bond with a B overhang at time t. At long times, when equilibrium is reached, p_b approaches the (concentration-dependent) value fixed by the binding equilibrium constant.[53]

Similar to other studies, we consider a cluster as an object formed by NSs joined together *via* AB bonds. Each cluster has at the most one unreacted A site. This A site can bind to any of the unreacted B sites belonging to the same cluster (hence forming an intracluster bond, also indicated as a loop) or to an unreacted B site of another cluster, thus merging the two clusters. The FS mean-field theory[5,54] provides a prediction for the number of clusters $N_c(n)$ of size n

$$N_c(n) = N(1 - p_b)F(n, p_b), \tag{1}$$

where

$$F(n, p_b) = \frac{[(f-1)n]!}{n![(f-2)n+1]!} \frac{p_b^{n-1}(f-1-p_b)^{(f-2)n+1}}{(f-1)^{(f-1)n}} \tag{2}$$

and the distribution is normalized such that

$$\sum_n nN_c(n) = N. \tag{3}$$

The FS theory, which is formally equivalent to a constrained maximization of the combinatorial entropy,[55] is based on the evaluation of the number of distinct cluster arrangements, with the restriction that the N monomers are connected by Np_b bonds to form polydisperse loopless clusters. In eqn (1), the term $N(1 - p_b)$ is equal to the total number of clusters and it clearly reveals the mean-field approximation, being the number of clusters equal to the number of particles minus the number of bonds (Np_b).

Fig. 2a and 3a show the cluster size distribution $N_c(n)/N$ at three different values of p_b (three different times during the simulation) for two different NS concentrations (c = 20 mg ml^{-1} and c = 2 mg ml^{-1}, respectively). The insets show the corresponding time evolution of p_b. As can be seen, the FS theoretical predictions, with no fit parameter, become incap-
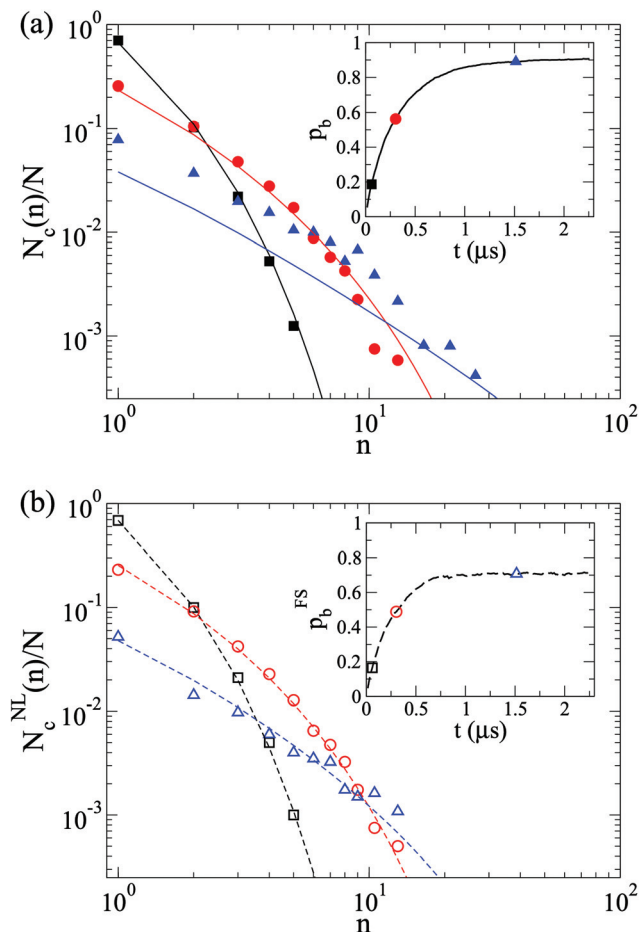
Fig. 2 The main graphs show the number distribution of clusters (symbols, panel a) and the number distribution of clusters in the FS subset (panel b) of size n for different simulation times (*i.e.* different bonding probabilities) for the simulation at $c = 20$ mg ml^{-1}, $N = 2000$. The lines are the theoretical predictions given by eqn (1) using p_b (panel a) and p_b^{FS} (panel b), respectively, whose evolution over the simulation time is displayed in the insets. Symbols and colors correspond to the time step relative to the curves in the main graphs. Legend: Black squares, $t \simeq 0.06$ µs; red circles, $t \simeq 0.30$ µs; blue triangles, $t \simeq 1.52$ µs.
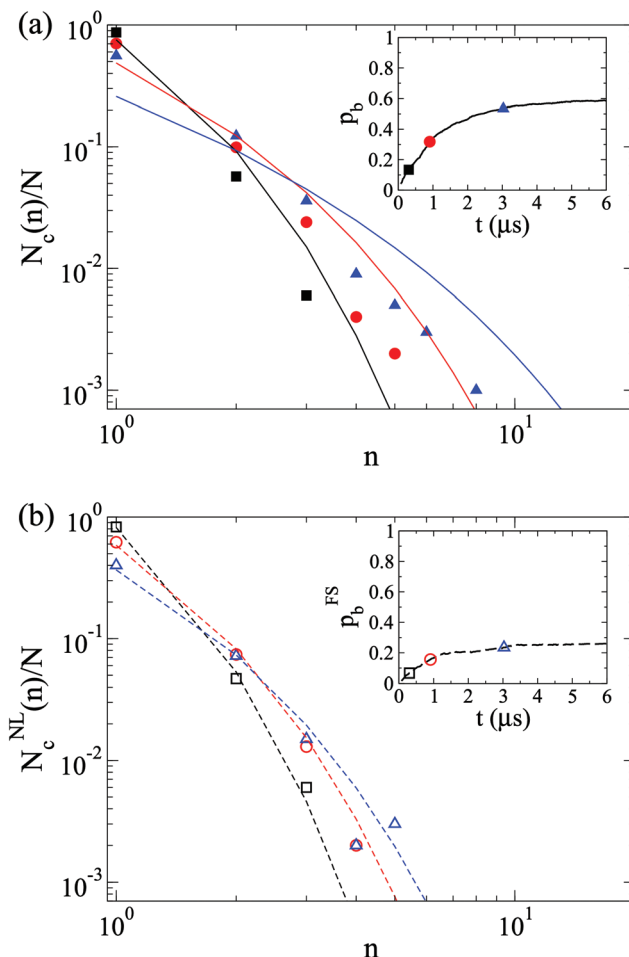
Fig. 3 Similar to Fig. 2, the main graphs show the number distribution of clusters (symbols, panel a) and the number distribution of clusters in the FS subset (panel b) for different simulation times for the simulation at $c = 2$ mg ml^{-1}, $N = 1000$. The lines are the theoretical predictions given by eqn (1) using p_b (panel a) and p_b^{FS} (panel b), respectively, whose evolution over the simulation time is displayed in the insets. Symbols and colors correspond to the time step relative to the curves in the main graphs. Legend: Black squares, $t \simeq 0.30$ µs; red circles, $t \simeq 0.91$ µs; blue triangles, $t \simeq 3.03$ µs.

able of representing the numerical data when $p_b \gtrsim 0.5$. This is very clear for the data at 20 mg ml^{-1}, for which the FS predictions underestimate the distribution of small clusters by more than a factor of two. The disagreement between theoretical predictions and numerical data at finite times could originate from kinetic pathways and/or due to the presence of closed loops, neglected in the FS approach. At long times, however, when thermodynamic equilibrium is approached, disagreement can only be ascribed to the presence of intracluster bonds.

To double-check the role of intracluster bonds, facilitated by the flexibility of the NS and by the growing density of B sites upon increasing the cluster size, we calculate the number of clusters with no reactive A sites, which is equal to the number of intracluster bonds. We also separate the clusters into two groups: the proper FS loopless clusters and the ones

with intracluster bonds. The latter violate the hypothesis of the FS theory, which assumes that each cluster has one and only one reactive A site. For the clusters with no loops (NL), we calculate the total number of existing bonds $\#_b^{NL}$ and the total number of particles composing these clusters N_{NL}. The ratio between these two numbers provides the bond probability p_b^{FS} = $\#_b^{NL}/N_{NL}$ for the subset of clusters satisfying the FS hypotheses. The size distributions of the FS-compatible NL clusters for different simulation times are shown in Fig. 2b and 3b and compared with the FS predictions from eqn (1), identifying N with N_{NL} and p_b with p_b^{FS}. The quality of the agreement confirms that for the NL clusters, where the FS hypotheses hold by default, the mean-field predictions properly represent the data, suggesting that the time evolution of the aggregation process is sufficiently well-described by the equilibrium solutions.[56]

Additionally, we provide a quantification of the structural properties of the clusters. We evaluate the mean squared radius of gyration

$$\langle R_g^2 \rangle(n) = \frac{1}{2n^2} \left\langle \sum_{j=1}^{n} \sum_{k=1}^{n} (\mathbf{r}_j - \mathbf{r}_k)^2 \right\rangle, \quad (4)$$

where $\mathbf{r}_j$ is the position of the center of mass of the j-th NS belonging to an aggregate of size n and the angular parentheses indicate an ensemble average over all the clusters with the same size n and over time. For fractal objects,

$$\langle R_g^2 \rangle(n) \sim n^{\frac{2}{d_f}}, \quad (5)$$

where d_f is the fractal exponent. A power law fit of the data presented in Fig. 4 shows that clusters larger than 3–4 monomers grow with $d_f \approx 2$.

Going back to the cluster size distribution, we next provide an extension of the FS theory to account for intracluster bonds. We recall that, formally, for weakly interacting clusters, the probability of formation of a cluster of size n in equilibrium is proportional[55,57] to its partition function Q_n multiplied by an (concentration-dependent) activity z^n. As shown in section D of the ESI,† the FS equation can indeed be recast in this ideal gas of clusters thermodynamic formalism as

$$N_c^{NL}(n) = Q_n^{NL} z^n, \quad (6)$$

where the partition function of a loopless cluster

$$Q_n^{NL} = \frac{V}{V_{ref}} \frac{[(f-1)n]!}{n![(f-2)n+1]!} \left(e^{-\beta F_{bond}} \right)^{n-1} \quad (7)$$

which is proportional to the system volume, measured in units of a reference volume V_{ref}, and is composed of a free-energy dependent term, $\exp[-\beta F_{bond}(V_{ref}, T)]$, modelling the contri-
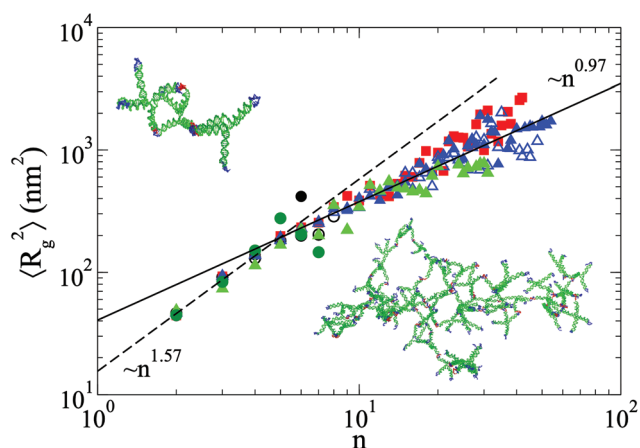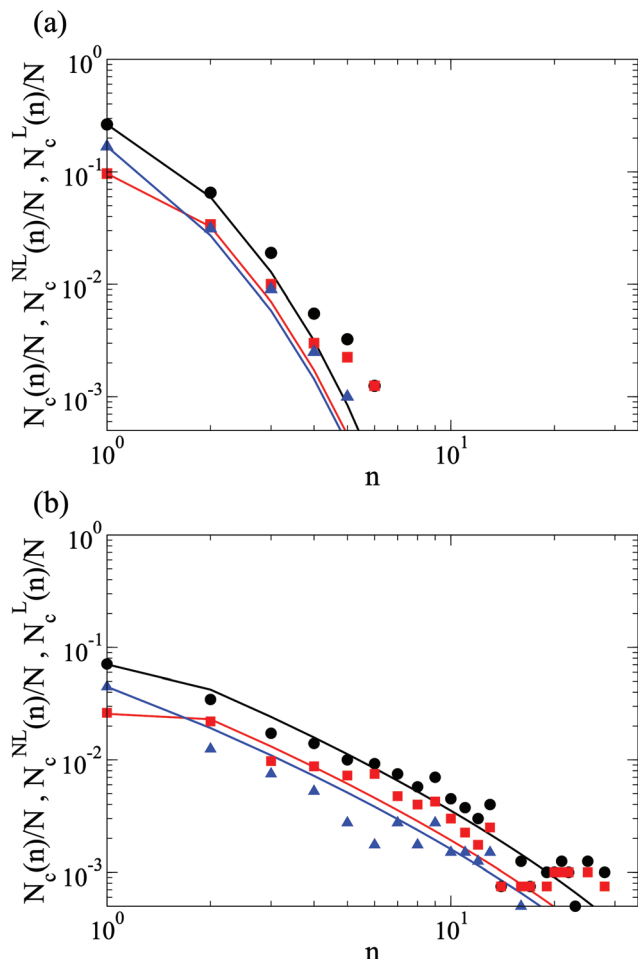

Fig. 4 Average squared radius of gyration as a function of the cluster size n for different simulations. Power law fits for $n \leq 5$ and $n > 5$ are represented by the dashed and full lines, respectively. Legend: Triangles, $c = 20$ mg ml^{-1} for $N = 300$ (green) and $N = 2000$ (blue, open and full); red squares, $c = 10$ mg ml^{-1}, $N = 2000$; circles, $c = 2$ mg ml^{-1} for $N = 300$ (dark green) and $N = 2000$ (black). In the top-left portion: a small cluster ($n = 3$). In the bottom-right portion: a large cluster ($n = 49$).

bution of the formation of $n - 1$ bonds, and the FS combinatorial entropic term. $\beta = 1/k_B T$ as usual. z plays the role of an activity and its value controls the concentration of the system. Being $Q_1^{NL} = V/V_{ref}$, it is possible to identify z with the nondimensional concentration of unbonded particles $N_c^{NL}(1)V_{ref}/V$. The mapping between p_b and F_{bond} is provided in section D of the ESI.†

To include the possibility of intracluster bonds, we sum the partition functions associated with the configurations without and with loops and we write the cluster size distribution of the entire system as

$$N_c(n) = (Q_n^{NL} + Q_n^L) z^n. \quad (8)$$

The partition function Q_n^L, compared to Q_n^{NL}, must include two terms: (i) an additional factor $\exp(-\beta F_{bond})$, which accounts for the presence of the intracluster extra bond, and (ii) a model-dependent factor $g(n,\beta)$, which quantifies the free-energy gain of forming an intracluster bond. The factor $g(n,\beta)$ includes the relative number of microscopic configurations with an intracluster loop (with respect to a loopless cluster). It may also include the free-energy cost of bringing the selected B site close to the unbonded A site, and thus it could, in principle, (weakly) depend on T if the monomer arms are not quite flexible. The partition function Q_n^L can then be written as

$$Q_n^L = g(n,\beta) Q_n^{NL} e^{-\beta F_{bond}}. \quad (9)$$

The unknown $g(n,\beta)$ can be estimated by evaluating the ratio between the number of clusters of size n with and without loops from the simulated configurations at long times, when equilibrium has been reached,

$$\frac{N_c^L(n)}{N_c^{NL}(n)} = \frac{Q_n^L}{Q_n^{NL}} = g(n,\beta) e^{-\beta F_{bond}}. \quad (10)$$

It is worth noting that the ratio $N_c^L(n)/N_c^{NL}(n)$ depends only on the temperature but not on the concentration.

Fig. 5 compares the predictions of eqn (8) and (9) with the numerical data from the simulations at $c = 2$ mg ml^{-1} (panel a) and $c = 20$ mg ml^{-1} (panel b) at long times, close to (if not at) equilibrium. In the comparison, $g(n,\beta)$ and $\exp(-\beta F_{bond})$ are the only fit parameters (β is fixed), identical for all the densities. The values of z are fixed by the concentration of loopless monomers. Despite the intrinsic noise of the data, the theoretical predictions well represent the numerical values at all densities. The fit suggests that $g(n,\beta)$ is essentially constant already for $n \gtrsim 2$ (*i.e.* for clusters composed of two monomers or more). This confirms that the unbonded A site essentially binds with a B site on the same particle or, at most, with one of its neighbor monomers. To provide additional support for this statement, we investigate the distribution of loop sizes,§ confirming that the average loop size is quite small ($\simeq 1.7 \pm 1$).

Even at the coarse-grained level of the oxDNA model, simulations are still too demanding to access lower temperatures

---

§ The loop size is defined as the number of monomers in a cluster that form a closed cycle of bonds.

(a)

(b)

Fig. 5 Comparison of the prediction of eqn (8) and (9) (lines) with the simulation equilibrium data (symbols). Data are obtained by averaging the cluster size distribution from two simulations at the same simulation time. Panel a refers to the simulation at $c = 2$ mg ml^{-1} ($t \simeq 6.9$ μs). Panel b refers to $c = 20$ mg ml^{-1} ($t \simeq 2.85$ μs). Legend: Circles, $N_c(n)/N$; squares, $N_c^L(n)/N$; triangles, $N_c^{NL}(n)/N$.

than the one we have studied. However, the previous model allows us to predict the expected cluster size distribution at low T, when the driving force for bonding becomes quite strong and all A sites have reacted. Under these conditions, Q^{NL} is negligible compared to Q^L for all n. The cluster size distribution will coincide with the distribution of the clusters with an intracluster bond and, therefore, will be given by

$$N_c(n) = \frac{V}{V_{ref}} g(n, \beta) \frac{[(f-1)n]!}{n![(f-2)n+1]!} \left(e^{-\beta F_{bond}}\right)^n z^n, \quad (11)$$

which can be recast in the form

$$N_c(n) = \frac{V}{V_{ref}} g(n, \beta) \frac{[(f-1)n]!}{n![(f-2)n+1]!} \left(\frac{V_{ref} N_c(1)}{V g(1, \beta)}\right)^n. \quad (12)$$

The term $[V_{ref} N_c(1)/V g(1,\beta)]$ acts as a renormalized activity. Its value can be tuned to fix the average concentration. The predicted low-T cluster size distributions for the three different
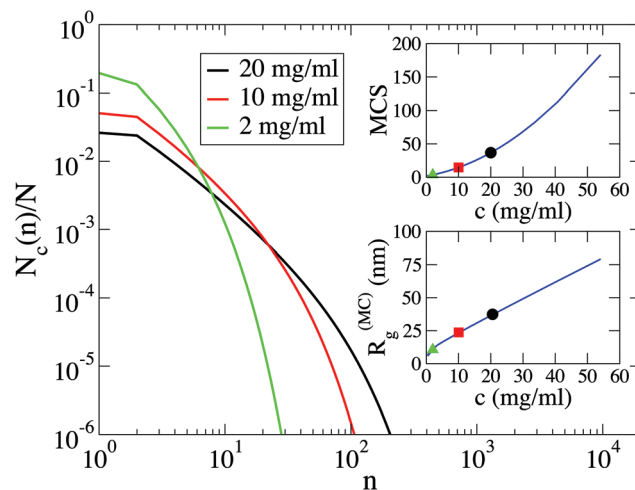
Fig. 6 Main: Predicted number distribution of clusters $N_c(n)/N$ in the limit of the fully bonded system (eqn (12)). Top inset: Mean cluster size $\left(\text{MCS} = \sum_n n^2 N_c(n) / \sum_n n N_c(n)\right)$ as a function of the NS concentration. The line is obtained by analytically computing the mean cluster size from the low temperature cluster distribution of eqn (12). Bottom inset: Radius of gyration of the mean cluster $R_g^{(MC)}$ as a function of the NS concentration. The line is obtained by combining the MCS dependence with the fit of the radius of gyration shown in Fig. 4.

investigated densities are shown in Fig. 6. The insets of the same figure show the associated mean cluster size (MCS) and the relative radius of gyration of the mean cluster $R_g^{(MC)}$ as a function of the NS concentration, respectively. From these results, we can formulate three important considerations. First, eqn (12) shows that the temperature (apart from the weak dependence entering in $g(n,\beta)$) does not play any role: once all possible bonds are formed, the equilibrium distributions are the ones that maximize the entropy. Second, the same equation shows that the NS concentration modulates the cluster size distribution, at odds with the FS predictions, which suggest the formation of an infinite percolating cluster incorporating all monomers. Third, and more importantly, the cluster size distribution remains finite at all physical values of the NS concentrations. Hence, the chance to form intracluster bonds eliminates the possibility to approach the percolation transition. In a more physical way, eqn (12) tells us that, when particles can satisfy all their bonds within the same cluster, the fully bonded (low-T) configuration is not the percolating one. Rather, the equilibrium low-T state exploits the entropic gain provided by the exploration of the system volume by a multiplicity of clusters, modulated by a slightly modified – by $g(n, \beta)$ – FS combinatorial term.

Experimental results

The simulation study has revealed that the presence of intracluster bonds strongly limits the formation of larger clusters in the system, preventing the possibility to reach the percolation point even when all possible bonds are formed ($p_b = 1$). In this case, it has also shown that the cluster size distribution

is strongly concentration-dependent, with a MCS (inset of Fig. 6) that is predicted to remain finite at all experimentally accessible NS concentrations.

To test the numerical findings, we realize the very same system in the laboratory and examine it *via* DLS. With experiments, we are not limited to the investigation of one single temperature. Instead, by changing T, we can probe different p_b values and even explore the T-window where all bonds are formed ($p_b = 1$ for $T \lesssim 20\ °C$), as shown by the melting profile of the sticky sequences reported in section A of the ESI.† In addition, experiments allow to probe the equilibrium properties of the system. Samples are left to equilibrate for several minutes (~40 min), a time sufficiently long to break and reform several bonds between the A and B sticky ends. We checked that all results are reproducible upon increasing and decreasing T scans and are not affected by aging nor by the previous history.

Fig. 7a–c show the autocorrelation functions of the scattered field $g_1(t)$ for the three investigated NS concentrations (c = 2 mg ml^{-1}, 10 mg ml^{-1}, and 20 mg ml^{-1}) and for the explored Ts. For all samples and temperatures, the correlation functions decay to zero within the experimentally accessible time window (10 s), confirming the sample ergodicity. For all studied concentrations, the system is far from a percolation transition, consistent with the numerical simulations and the proposed theoretical extension of the FS theory.

To quantify the slowing down of the dynamics and to extract a typical (slow) relaxation time, the correlation curves are fitted to a double stretched exponential function (see section E of the ESI† for comparison to the fit function)

$$g_1(t) = (1 - A_s)\exp(-t/\tau_f) + A_s \exp(-t/\tau_s)^{\beta_s}, \quad (13)$$

where τ_f and τ_s are the relaxation times of the fast and slow relaxation processes, respectively, A_s is the amplitude of the slow process, and β_s its stretching exponent. The slow relaxation time is better represented by its average value, defined as

$$\langle \tau_s \rangle = \frac{\int_0^\infty t \exp(-t/\tau_s)^{\beta_s}dt}{\int_0^\infty \exp(-t/\tau_s)^{\beta_s}dt} = \frac{\tau_s}{\beta_s}\Gamma\left(\frac{1}{\beta_s}\right), \quad (14)$$

where Γ is the gamma function.

The insets of Fig. 7a–c show the values of β_s, which are associated with the slow relaxation process. For all the measurements, the values lie within the range $0.4 \lesssim \beta_s \lesssim 0.6$. The values of the slow relaxation time are shown in Fig. 8 for all the concentrations and temperatures. To eliminate the trivial effect of the temperature dependence of the solvent viscosity, the times are rescaled to the viscosity η_{solv} of the 250 mM NaCl solvent at the highest investigated temperature ($T_{ref} \simeq 48.5\ °C$) as

$$\tau_s^*(T) = \langle \tau_s \rangle(T)\frac{\eta_{solv}(T_{ref})}{\eta_{solv}(T)}. \quad (15)$$

In general, two distinct processes – which can only be disentangled in particular conditions – contribute to the relax-
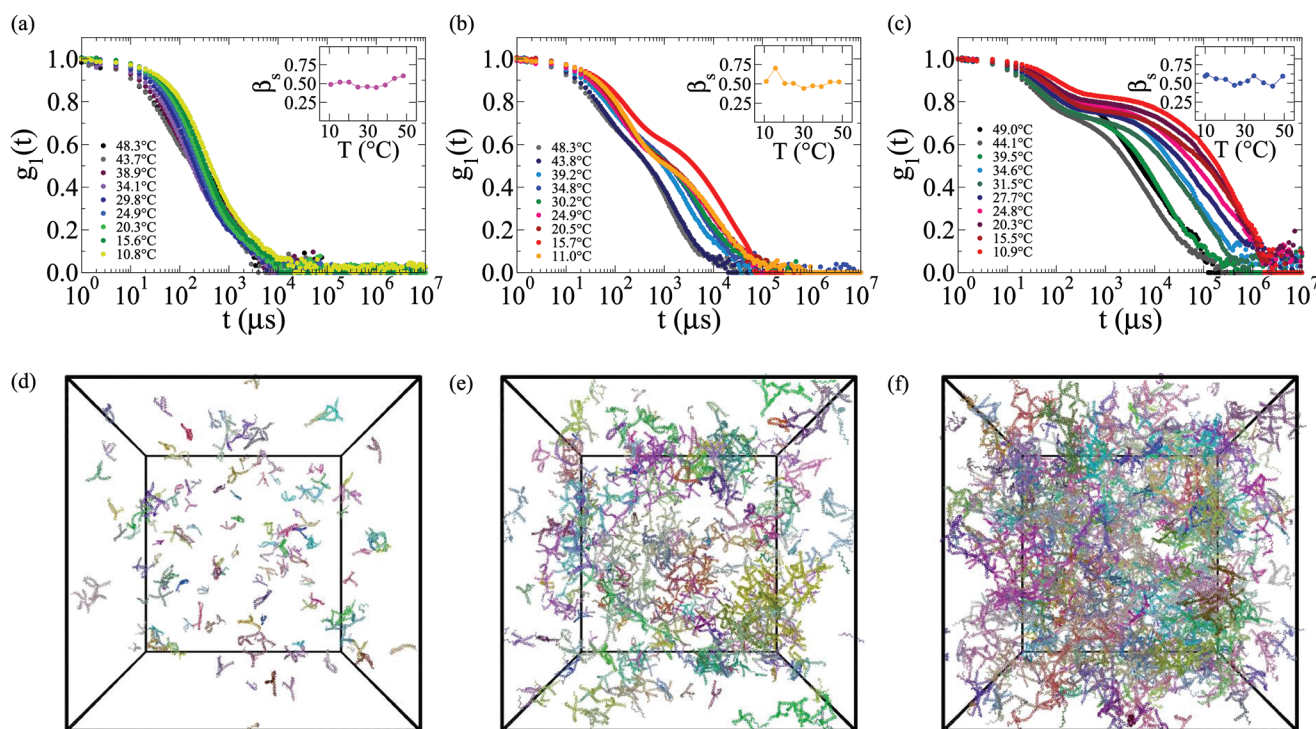
Fig. 7 Upper panels (a–c): DLS results ($g_1(t)$) showing the T-behavior of the system for the three different investigated NS concentrations: (a) 2 mg ml^{-1}, (b) 10 mg ml^{-1}, and (c) 20 mg ml^{-1}. Lower panels (d–f): Snapshots obtained from simulations performed at T = 45 °C at the same concentrations of panels a–c. The boxes are displayed on the same scale (box side is $L \simeq 220$ nm). Different clusters are indicated with different colors.
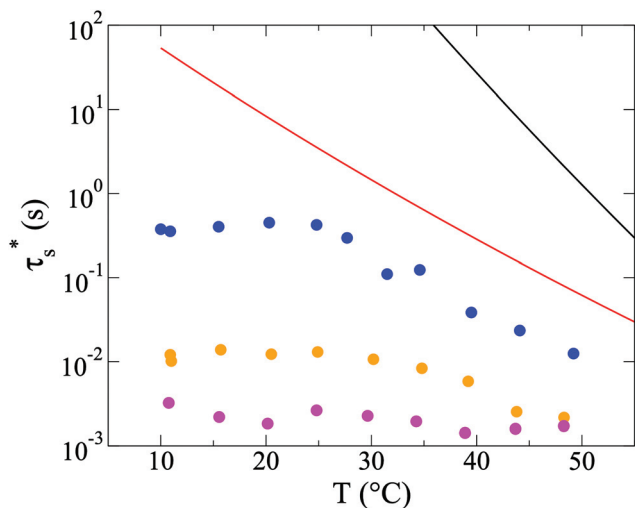
Fig. 8 T-Dependence of the (viscosity rescaled) slow decay time τ_s^* for samples at the three different concentrations: 20 mg ml^{-1} (blue), 10 mg ml^{-1} (orange), and 2 mg ml^{-1} (magenta). The lines are the two estimated bond lifetimes, calculated according to eqn (16) using $\alpha = 0.5$ (red) and $\alpha = 1$ (black), to provide support to the assumption that the relaxation process takes place in a fixed bonding pattern.

ation process probed by DLS: cluster restructuring and cluster diffusion.

The cluster restructuring times, *i.e.* the times associated with the cluster fracturing and/or the coalescence of different aggregates, are related to the breaking of bonds. Hence, the bond-breaking times control the cluster restructuring. In our DNA design, explicitly selected to generate a fully bonded system for $T \lesssim 20$ °C, 9 base pairs take part to form the AB bond, resulting in an enthalpic ΔH and entropic ΔS contribution to the binding free-energy ΔG equal to $\Delta H = -61.5$ kcal mol^{-1} and $\Delta S = -178.3$ cal mol^{-1} K^{-1}, evaluated using the web application DINAMelt[58] based on SantaLucia.[53] A reasonable estimate of the bond lifetime is[41]

$$\tau = \tau_0 e^{\alpha \Delta G / RT}, \quad (16)$$

where τ_0 is of the order of a few ms and $R = 1.987$ cal mol^{-1} K^{-1}. The value of the coefficient α is between 0.5 and 2, depending on the degree of cooperativity of the bond-breaking process.[59] Here, we use $\tau_0 = 3$ ms, as found in ref. 41. In Fig. 8, we show the expected T-dependence of the bond lifetime as predicted by eqn (16) for $\alpha = 0.5$ and $\alpha = 1$. In both cases, the experimental times are faster than the bond-breaking time, even when the most conservative estimate of the bond lifetime is selected ($\alpha = 0.5$). It is known, however, that the estimates of hybridization/de-hybridization kinetic rates are approximate. Therefore, one could wonder whether the slow characteristic relaxation times, measured at high temperatures, might be associated with the bond breaking between NSs. However, in such case, the measured slow relaxation times should be very similar for all the concentrations. In fact, the bond-breaking time is mostly controlled by the hybridization of (identical) sticky binding sequences (apart from a small logarithmic cor-

rection depending on c). This implies that the cluster diffusion is the main mechanism for the decorrelation of the density fluctuations in this system. Additionally, we observe a crossover between a high-T regime, where τ increases on lowering the temperature, and a low-T regime, where τ is constant. The crossover temperature is very close to the one where the theoretically predicted bond probability approaches unity (see Fig. S1 of the ESI†). This suggests that the temperature interval ($T \lesssim 20$ °C) where the relaxation time is constant coincides with the regime where all AB bonds are formed, while the increase upon temperature reduction at high T is linked to the progressive growth of the average cluster size.

At the lowest density ($c = 2$ mg ml^{-1}, Fig. 7a), where the hypothesis of independent clusters is more realistic, the relaxation process is quite fast and it does only show a weak hint of two-step relaxation, suggesting that the experiments are probing the free diffusion of limited-size clusters. In Fig. S3d of the ESI,† we compare the experimental data at $c = 2$ mg ml^{-1} with both the simple and double stretched exponential (eqn (13)) fits. We also notice that, even at low c, the sample is not monodisperse but composed of small aggregates (with a polydispersity lower than the one for the larger concentrations). The relaxation time grows only by a factor of ≈ 5 when T varies from 50 °C to 10 °C (corresponding to p_b from ≈ 0 to ≈ 1). At $c = 10$ mg ml^{-1} (Fig. 7b), a weak additional relaxation process appears, signalling the onset of interactions between clusters. The respective image from the simulation (Fig. 7e), which we can use as a visual aid to interpret the DLS data, suggests that it may become appropriate to tentatively interpret the fast decay as originating from the cluster-free diffusion, while the slow one comes from the confinement induced by the presence of nearby clusters. This hypothesis is supported by the data at $c = 20$ mg ml^{-1} (Fig. 7c) – and again by the simulation snapshot in Fig. 7f –, where the excluded-volume interactions between clusters are significant, as illustrated by the pronounced two-step decay of the correlation functions.

We then intend to compare the simulation results with the experiments for the sample at $c = 2$ mg ml^{-1}. We evaluate the intensity-averaged¶ hydrodynamic radius from the cluster size distribution obtained from the simulations at $T = 45$ °C (*cf.* Fig. 5) and from the low-T theory (*cf.* Fig. 6). For each cluster of size n, we calculate the hydrodynamic radius following the recipe described in ref. 60, which provides the equivalent ellipsoid associated with the cluster. The actual hydrodynamic radius is then computed from the known formulae.[61] Finally, as done for the radius of gyration, we evaluate the cluster size dependence of the ensemble average of R_h. The power law fit to the data gives $\langle R_h \rangle (n) \simeq R_0 n^{\gamma'}$ with

$$R_0 \simeq \begin{cases} 5.0 \text{ nm} & \text{if } n \le 5 \\ 6.5 \text{ nm} & \text{if } n > 5 \end{cases} \quad (17)$$

¶The intensity average for the simulation data is obtained by performing the average weighted by the distribution $n^2 N_c(n)$.

and

$$\gamma = \begin{cases} 0.68 & \text{if } n \leq 5 \\ 0.53 & \text{if } n > 5 \end{cases}. \qquad (18)$$

For further details, see section F of the ESI.† At high temperature, equal to the one numerically investigated (T = 45 °C), we obtain from the simulations an intensity-averaged hydrodynamic radius of $\langle R_h \rangle_z \simeq 9.3$ nm (using the radius of gyration, one obtains $\langle R_g \rangle_z \simeq 8.1$ nm). By analyzing the correlation time associated with the fast relaxation τ_f, we estimate a mean hydrodynamic radius $R_h \simeq 5.5$ nm. At low temperature, the numerical prediction (see Fig. 6) provides an expected radius $\langle R_h \rangle_z \simeq 14.9$ nm ($\langle R_g \rangle_z \simeq 13.2$ nm), to be compared to the experimental value of $R_h \simeq 22.5$ nm.

The discrepancy between the experimental values and the ones estimated from the simulations – which is about 50% at low temperature and grows to a factor of ∼0.7 at T = 45 °C – can be mainly ascribed to the approximate model we used for the calculation of the hydrodynamic radius. Another possible source of error can be related to the numerical noise of the data used in the fit to obtain $g(n,\beta)$.

Conclusions

This article, based on a combined numerical, theoretical, and experimental study, reports numerous relevant findings.

• It shows that DNA-made NSs with precise binding topologies can be nowadays produced in bulk quantities and exploited as highly precise model systems to tackle the physics of aggregating particles, either directional colloids or functional polymers. Here, we studied the aggregation behavior of AB$_3$ monomers made of DNA NSs with the aim of clarifying the behavior of hyperbranched aggregation when intracluster bonds are present.

• It reports an extensive MD simulations study, based on a high-quality coarse-grained potential for the DNA interactions, to investigate the equilibrium cluster size distribution at a fixed temperature and at different initial monomer concentrations. We found that the mean cluster size increases with the particle concentration, from a suspension of monomers and small clusters, at low c, to a highly polydisperse solution at large concentrations. Simulations also suggested that the presence of intracluster bonds eliminates the possibility to reach the percolation transition.

• It reports a novel theoretical way to include the presence of intracluster bonds in the formalism. First, the FS approach is translated in an ideal gas of clusters formalism, which is then extended to include also the clusters with intracluster bonds. Including a model-dependent (but concentration-independent) quantity $g(n,\beta)$ – which can be determined as a fit parameter – it becomes possible to predict the cluster size distribution without limiting assumptions. The theory explains why percolation is avoided when intracluster bonds are possible. Comparison with MD simulation data supports the quality of the modelling.

• It reports DLS measurements of the same system studied numerically and theoretically to provide evidence that, at odds with the FS predictions, percolation is not encountered in this system. Indeed, by decreasing the temperature, and exploiting the reversibility of the sticky-end pairing, it is possible to investigate the effect of bonding, up to the point where all bonds in the system are formed. The data we collected strongly support the idea that the polydispersity of the aggregates remains finite at low temperatures. Additionally, it shows that the low-T equilibrium cluster size distribution only depends on the initial monomer concentration.

We believe that these results will be also valuable for the community interested in the biotechnological applications of hyperbranched polymers, due to the biocompatibility and versatility of DNA as a building block for complex and innovative materials.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 E. Bianchi, P. Tartaglia, E. L. Nave and F. Sciortino, *J. Phys. Chem. B*, 2007, **111**, 11765.
2 S. Corezzi, C. D. Michele, E. Zaccarelli, D. Fioretto and F. Sciortino, *Soft Matter*, 2008, **4**, 1173.
3 M. Wertheim, *J. Stat. Phys.*, 1984, **35**, 19–34.
4 M. Wertheim, *J. Stat. Phys.*, 1986, **42**, 459–476.
5 P. Flory, *J. Am. Chem. Soc.*, 1941, **63**, 3083.
6 W. Stockmayer, *J. Chem. Phys.*, 1943, **11**, 45–55.
7 E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli and F. Sciortino, *Phys. Rev. Lett.*, 2006, **97**, 168301.
8 F. Sciortino, E. Bianchi, J. Douglas and P. Tartaglia, *J. Chem. Phys.*, 2007, **126**, 194903.
9 Y. Kim and O. Webster, *J. Am. Chem. Soc.*, 1990, **112**, 4592.
10 Y. Kim and O. Webster, *Macromolecules*, 1992, **25**, 5561.
11 T. Cuneo and H. Gao, *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.*, 2020, e1640.
12 C. Jochum, N. Adžić, E. Stiakakis, T. Derrien, D. Luo, G. Kahl and C. Likos, *Nanoscale*, 2019, **11**, 1604–1617.
13 B. Zhou, W. Huang, J. Liu, X. Zhu and D. Yan, *Adv. Mater.*, 2010, **22**, 4567.
14 J. Liu, W. Huang, Y. Pang and D. Yan, *Chem. Soc. Rev.*, 2015, **44**, 3942.
15 R. van Benthem, *Prog. Org. Coat.*, 2000, **40**, 203–214.

16 A. Zotti, S. Zuppolini, A. Borriello and M. Zarrelli, *Nanomaterials*, 2019, **9**, 418.

17 C. Paleos, D. Tsiourvas, Z. Sideratou and L.-A. Tziveleka, *Expert Opin. Drug Delivery*, 2010, **7**, 1387.

18 V. Gajbhiye, L. Escalante, G. Chen, A. Laperle, Q. Zheng, B. Steyer, S. Gong and K. Saha, *Nanoscale*, 2013, **6**, 521–531.

19 M. Qi, S. Duan, B. Yu, H. Yao, W. Tian and F. Xu, *Polym. Chem.*, 2016, **7**, 4334.

20 L. Zhao, X. Wu, X. Wang, C. Duan, H. Wang, A. Punjabi, Y. Zhao, Y. Zhang, Z. Xu, H. Gao and G. Han, *ACS Macro Lett.*, 2017, **6**, 700–704.

21 M. Qi, S. Duan, B. Yu, H. Yao, W. Tian and F. Xu, *Colloids Surf., B*, 2019, **182**, 110375.

22 G. Feng, D. Ding and B. Liu, *Nanoscale*, 2012, **4**, 6150.

23 A. M. Malekzadeh, A. Ramazani, S. T. Rezaei and H. Niknejad, *J. Colloid Interface Sci.*, 2017, **490**, 64–73.

24 C. Pitois, D. Wiesmann, M. Lindgren and A. Hult, *Adv. Mater.*, 2001, **13**, 1483.

25 H. Shi, X. Chen, S. Liu, H. Xu, Z. An, L. Ouyang, Z. Tu, Q. Zhao, Q. Fan, L. Wang and W. Huang, *ACS Appl. Mater. Interfaces*, 2013, **5**, 4562.

26 X. Liu, H.-J. Liu, F. Cheng and Y. Chen, *Nanoscale*, 2014, **6**, 7453.

27 Y. Zheng, H. Cao, B. Newland, Y. Dong, A. Pandit and W. Wang, *J. Am. Chem. Soc.*, 2011, **133**, 13130.

28 J. Lyu, Y. Gao, Z. Zhang, U. Greiser, H. Tai and W. Wang, *Sci. China: Chem.*, 2018, **61**, 319–327.

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

30 S. Biffi, R. Cerbino, G. Nava, F. Bomboi, F. Sciortino and T. Bellini, *Soft Matter*, 2015, **11**, 3132.

31 L. Rovigatti, F. Bomboi and F. Sciortino, *J. Chem. Phys.*, 2014, **140**, 154903.

32 N. Conrad, T. Kennedy, D. Fygenson and O. Saleh, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 7238.

33 T. Ouldridge, A. Louis and J. Doye, *J. Chem. Phys.*, 2011, **134**, 085101.

34 B. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. Schreck, F. Romano, T. Ouldridge, R. Tsukanov, E. Nir, A. Louis and J. Doye, *J. Chem. Phys.*, 2015, **142**, 234901.

35 N. Seeman, *Nature*, 2003, **421**, 427–431.

36 N. Seeman, *NanoBiotechnology Protocols*, Springer, 2005, pp. 143–166.

37 T. Bellini, R. Cerbino and G. Zanchetta, *Liquid Crystals*, Springer, 2011, pp. 225–279.

38 M. Salamonczyk, J. Zhang, G. Portale, C. Zhu, E. Kentzinger, J. Gleeson, A. Jakli, C. D. Michele, J. Dhont, S. Sprunt and E. Stiakakis, *Nat. Commun.*, 2016, **7**, 13358.

39 J. Fernandez-Castanon, F. Bomboi, L. Rovigatti, M. Zanatta, A. Paciaroni, L. Comez, L. Porcar, C. Jafta, G. Fadda, T. Bellini and F. Sciortino, *J. Chem. Phys.*, 2016, **145**, 084910.

40 D. Nguyen and O. Saleh, *Soft Matter*, 2017, **13**, 5421.

41 F. Bomboi, S. Biffi, R. Cerbino, T. Bellini, F. Bordi and F. Sciortino, *Eur. Phys. J. E*, 2015, **38**, 64.

42 Z. Xing, A. Caciagli, T. Cao, I. Stoev, M. Zupkauskas, T. O'Neill, T. Wenzel, R. Lamboll, D. Liu and E. Eiser, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 8137.

43 I. Stoev, T. Cao, A. Caciagli, J. Yu, C. Ness, R. Liu, R. Ghosh, T. O'Neill, D. Liu and E. Eiser, *Soft Matter*, 2020, **16**, 990–1001.

44 F. Spinozzi, M. Ortore, G. Nava, F. Bomboi, F. Carducci, H. Amenitsch, T. Bellini, F. Sciortino and P. Mariani, *Langmuir*, 2020, **36**, 10387.

45 J. Fernandez-Castanon, M. Zanatta, L. Comez, A. Paciaroni, A. Radulescu and F. Sciortino, *ACS Macro Lett.*, 2019, **8**, 84–87.

46 G. Nava, T. Yang, V. Vitali, P. Minzioni, I. Cristiani, F. Bragheri, R. Osellame, L. Bethge, S. Klussmann, E. Paraboschi, R. Asselta and T. Bellini, *Soft Matter*, 2018, **14**, 3288.

47 R. Brady, W. Kaufhold, N. Brooks, V. Foderà and L. D. Michele, *J. Phys.: Condens. Matter*, 2019, **31**, 074003.

48 J. SantaLucia, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 1460.

49 J. A. Holbrook, M. W. Capp, R. M. Saecker and M. T. Record, *Biochemistry*, 1999, **38**, 8409.

50 https://dna.physics.ox.ac.uk.

51 J. Russo, P. Tartaglia and F. Sciortino, *J. Chem. Phys.*, 2009, **131**, 014504.

52 B. Berne and R. Pecora, *Dynamic light scattering: with applications to chemistry, biology, and physics*, Courier Corporation, 2000.

53 J. SantaLucia, *Nucleic Acid Res.*, 1996, **28**, 1929.

54 M. Rubinstein and R. Colby, *Polymer Physics*, OUP Oxford, 2003.

55 F. Sciortino, *Riv. Nuovo Cimento Soc. Ital. Fis.*, 2019, **42**, 511–548.

56 S. Corezzi, D. Fioretto and F. Sciortino, *Soft Matter*, 2012, **8**, 11207.

57 T. Hill, *Statistical Mechanics: Principles and Selected Applications*, Dover Publications, 1987.

58 N. Markham and M. Zuker, *Nucleic Acid Res.*, 2005, **33**, W577.

59 F. Smallenburg and F. Sciortino, *Nat. Phys.*, 2013, **9**, 554–558.

60 L. Rovigatti, N. Gnan, A. Ninarello and E. Zaccarelli, *Macromolecules*, 2019, **52**, 4895.

61 J. Hubbard and J. Douglas, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1993, **47**, R2983.

62 https://sourceforge.net/projects/cogli1/.

Supplementary Material to *Hyperbranched DNA clusters*
E. Lattuada, D. Caprara, V. Lamberti, and F. Sciortino

A. Nupack DNA analysis

To confirm the thermodynamic binding behavior of the designed sequences, we use the NUPACK oligo simulator [1]. Based on SantaLucia thermodynamics calculations, NUPACK provides the melting profile for arbitrary sequences of oligomers at desired concentrations and salt conditions (more precisely, the T-dependence of the fraction of unbonded base pairs).

Fig. S1, shows the T-behavior associated with the self-assembly of the tetravalent particles as well as of the isolated sticky sequences (forming the AB bonds). Here, the strand concentrations are fixed at $313\,\mu$M (corresponding to the largest sample concentration $c = 20\,$mg/ml) and the salt concentration at $250\,$mM of NaCl, respectively. As can be noted, the gap in the melting temperatures between the particle assembly (see magenta points) with respect to their sticky-end hybridization (blue) guarantees a net separation between the self-assembly of the nanostructures ($T_{\mathrm{NS}} \simeq 77°$C) and the formation of the interparticle bonds ($T_b \simeq 42°$C).
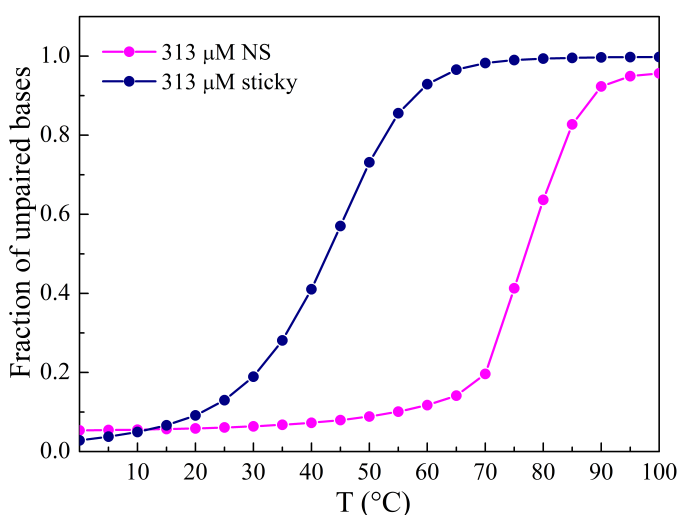

Figure S1: Melting curves calculated using NUPACK oligo analyzer [1] for both the NS arms (magenta) and the sticky tips (blue). Note that the self-assembly of the NS precedes on cooling the binding of the sticky sequences.

B. Preparation of simulation initial configuration

The simulation box is filled by placing the tetramers one after the other until the desired number N of tetramers is reached. Each tetramer is placed with random orientation and at a random position within the box. Then, the insertion energy is calculated: if the computed energy is larger than a predefined threshold (the thermal energy at $3 \times 10^4\,$K), the tetramer is moved to a new position/orientation. The energy check is then repeated until the insertion energy is less than the threshold. The resulting configuration is then equilibrated to $75°$C. This configuration – in which all tetramers are not bonded – is then used as the starting configuration at the desired simulated temperature.

C. Bond definition in the numerical study

In the numerical study, we have defined a bond between an A and a B arm as a double-helix section of at least five bases. This value is sufficiently large to exclude temporary association with just a few bases. To provide evidence that results are robust respect to this threshold value, we show in Fig. S2 the cluster distribution obtained at long times for the simulations at $c = 2\,$mg/ml, $N = 1000$ (left) and $c = 20\,$mg/ml, $N = 2000$ (right) when varying the number of bases used in the bond definition.
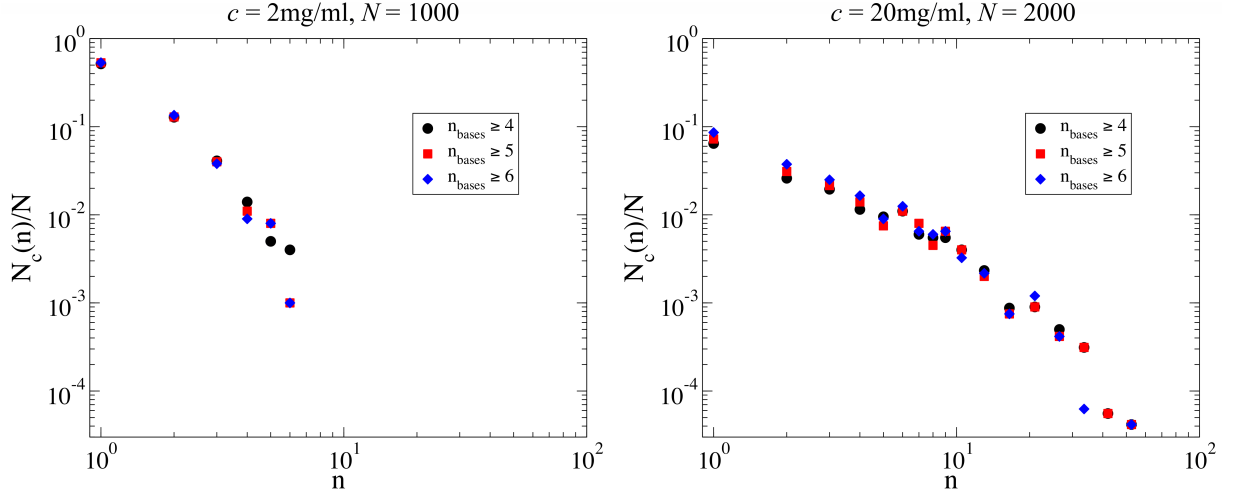
Figure S2: Effect of the choice of the minimum number of bases (n_{bases}) used to discriminate bonded and non-bonded nanostars.

D. Considerations about the cluster free-energy and Flory-Stockmayer approach

Here, we relate the bond probability assumed in the Flory-Stokmayer (FS) theory [2, 3] to the partition function of a bond. The FS theory assumes a value for the bond probability p_b. In the hyperbranched AB_{f-1} case, $p_b = \#_b/N$, being $\#_b$ the number of bonds in the system and N the total number of possible bonds (equal to the number of monomers). In a thermodynamic approach, p_b is controlled by the interaction potential and by the state variables T and V.

Let us consider a solution of N A-patches and $N(f-1)$ B-patches. We assume that an equilibrium is established between bonded and non-bonded interactions, the latter quantified by a bonding volume V_b and a bonding energy ϵ_{AB}. The mass-action law relates the number of reacted (N_{AB}) and unreacted (N_A and N_B) A- and B-patches to their partition functions as [4]

$$\frac{N_{AB}}{N_A N_B} = \frac{Q_{AB}}{Q_A Q_B}. \tag{1}$$

Neglecting intramolecular effects, to a first approximation, $Q_A = Q_B = V/\Lambda^3$ (where Λ is the de Broglie wavelength originating from the integration over momenta) and

$$Q_{AB} = \frac{V V_b}{\Lambda^6} e^{-\beta \epsilon_{AB}}, \tag{2}$$

so that

$$\frac{N_{AB}}{N_A N_B} = \frac{V_b}{V} e^{-\beta \epsilon_{AB}}. \tag{3}$$

Since N_{AB} is equal to the number of AB bonds in the system ($\#_b$), we can write for the number of unreacted A- and B-patches, respectively, $N_A = N - \#_b$ and $N_B = N(f-1) - \#_b$. Hence, the left hand side of Eq. (1) can be rewritten as

$$\begin{aligned} \frac{N_{AB}}{N_A N_B} &= \frac{\#_b}{(N - \#_b)[N(f-1) - \#_b]} \\ &= \frac{p_b}{(1 - p_b)[(f-1) - p_b]}, \end{aligned} \tag{4}$$

from which it follows that

$$\frac{p_b}{(1 - p_b)[(f-1) - p_b]} = \frac{N V_b}{V} e^{-\beta \epsilon_{AB}}. \tag{5}$$

This last relation provides the connection between the model parameters (V_b, ϵ_{AB}), V, and T and the bond probability p_b.

Starting from the thermodynamics expression for non-interacting clusters, one can write the number of clusters of size n resulting from the association process of the A- and B-patches as

$$N_c(n) = Q_n z^n, \tag{6}$$

where z plays the role of the Lagrange multiplier controlling the total number of particles in the system, which can also be expressed in terms of the number of unreacted particles (clusters of size 1, $N_c(1)$), since

$$z = \frac{N_c(1)}{Q_1} = \frac{N_c(1)\Lambda^3}{V}. \tag{7}$$

The number of free monomers can be estimated using the FS relation

$$N_c(1) = N(1 - p_b)\left(1 - \frac{p_b}{f - 1}\right)^{f-1}, \tag{8}$$

which expresses the fact that all the monomers must have one A and $(f - 1)$ B sites unbonded (we recall that the probability that a B site is unbonded is $\#_b/[N(f-1)]$). Then,

$$\frac{N_c(1)}{V} = \rho(1 - p_b)\left(1 - \frac{p_b}{f - 1}\right)^{f-1} \tag{9}$$

and we can write, by defining $\rho = N/V$ and $\rho_1 = N_c(1)/V$, and equating Eq. (7) and Eq. (9),

$$\frac{\rho_1}{\rho} = \frac{z\Lambda^{-3}}{\rho} = (1 - p_b)\left(1 - \frac{p_b}{f - 1}\right)^{f-1}. \tag{10}$$

Substituting this expression (to the power n) in the FS cluster size distribution (Eq. (1) of the manuscript, here reproduced)

$$N_c(n) = N(1 - p_b)\frac{[(f - 1)n]!}{n![(f - 2)n + 1]!}\frac{p_b^{n-1}(f - 1 - p_b)^{(f-2)n+1}}{(f - 1)^{(f-1)n}}, \tag{11}$$

one finds

$$N_c(n) = N(1 - p_b)D(n, f)\,p_b^{n-1}\frac{(f - 1 - p_b)^{1-n}}{(1 - p_b)^n}\left(\frac{z\Lambda^{-3}}{\rho}\right)^n, \tag{12}$$

where

$$D(n, f) = \frac{[(f - 1)n]!}{n![(f - 2)n + 1]!}. \tag{13}$$

Simplifying and making use of the relation in Eq. (5),

$$\frac{N_c(n)}{V} = \frac{N}{V}D(n, f)\left(\frac{NV_b}{V}e^{-\beta\epsilon_{AB}}\right)^{n-1}\left(\frac{z\Lambda^{-3}}{\rho}\right)^n$$
$$= D(n, f)\left(V_b e^{-\beta\epsilon_{AB}}\right)^{n-1}(z\Lambda^{-3})^n. \tag{14}$$

Defining a reference volume V_{ref}, one can define a reference bonding free-energy $\mathcal{F}_{\text{bond}}$ as [5]

$$e^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)} = \frac{V_b}{V_{\text{ref}}}e^{-\beta\epsilon_{AB}}, \tag{15}$$

such that

$$N_c(n) = \frac{V}{V_{\text{ref}}}D(n, f)\left(e^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)}\right)^{n-1}(zV_{\text{ref}}\Lambda^{-3})^n. \tag{16}$$

By comparing the definition of $\mathcal{F}_{\text{bond}}$ and Eq. (5), it follows that

$$\frac{p_b}{(1-p_b)[(f-1)-p_b]} = \frac{NV_{\text{ref}}}{V}\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}}, \tag{17}$$

which provides the link between the bond probability and the bonding free-energy.

Redefining z as $(zV_{\text{ref}}\Lambda^{-3})$ and comparing Eq. (16) and Eq. (6), the partition function of a cluster of size n with no loops (NL) can be identified as

$$Q_n^{\text{NL}} = \frac{V}{V_{\text{ref}}}D(n,f)\left(\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)}\right)^{n-1}, \tag{18}$$

which clearly shows the cluster center of mass contribution V, the combinatorial contribution $D(n,f)$, and the contribution arising from the $n-1$ bonds.

Assuming that loops are also possible, one needs to add the contribution which includes the loops to the FS partition function. The additional intracluster bond adds a term proportional to $V_b\mathrm{e}^{-\beta\epsilon_{AB}}$. Then, we propose to write the partition function of clusters with loops (L) as

$$Q_n^{\text{L}} = Q_n^{\text{NL}}\,g(n,\beta)\,\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)}, \tag{19}$$

where the system-dependent factor $g(n,\beta)$ accounts for the additional free-energy change. Specifically, it includes any combinatorial and any elastic free-energy contributions not accounted by the $V_b\mathrm{e}^{-\beta\epsilon_{AB}}$ term. Hence, the resulting cluster size distribution is

$$N_c(n) = N_{\text{NL}}(n) + N_{\text{L}}(n) \tag{20}$$
$$= \frac{V}{V_{\text{ref}}}D(n,f)\left(\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)}\right)^{n-1}\left(1+g(n,\beta)\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)}\right)z^n,$$

where $g(n,\beta)$ can be evaluated by calculating the ratio between the number of clusters of size n with and without loops, $N_{\text{L}}(n)/N_{\text{NL}}(n)$.

Note that, at low T, when all bonds are formed, $g(n,\beta)\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)} \gg 1$ and only the contribution from the clusters with a loop to the partition function survives and

$$Q_n = \frac{V}{V_{\text{ref}}}D(n,f)\left(\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)}\right)^n g(n,\beta) \tag{21}$$

and

$$N_c(n) = \frac{V}{V_{\text{ref}}}D(n,f)\,g(n,\beta)\left(z\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)}\right)^n. \tag{22}$$

This time

$$N_c(1) = \frac{V}{V_{\text{ref}}}g(1,\beta)\left(z\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)}\right), \tag{23}$$

so that

$$z\mathrm{e}^{-\beta\mathcal{F}_{\text{bond}}(V_{\text{ref}},T)} = \frac{\rho_1}{g(1,\beta)}, \tag{24}$$

where ρ_1 is the monomer's (unbonded particles) density when the volume is measured in units of V_{ref}, and

$$N_c(n) = D(n,f)\,g(n,\beta)\left(\frac{\rho_1}{g(1,\beta)}\right)^n, \tag{25}$$

which, neglecting the weak dependence on T of $g(1,\beta)$, does not depend on T any longer. The cluster size distribution has reached its "ground state" limit.

E. DLS data fitting

Fig. S3 compares the experimental data for the autocorrelation functions $g_1(t)$ with the fit performed using the double stretched exponential function of Eq. (13) in the main document at high (a), intermediate (b), and low temperatures (c). In panel d, we compare the fit using a simple exponential function with the experimental data and the double stretched exponential fit.
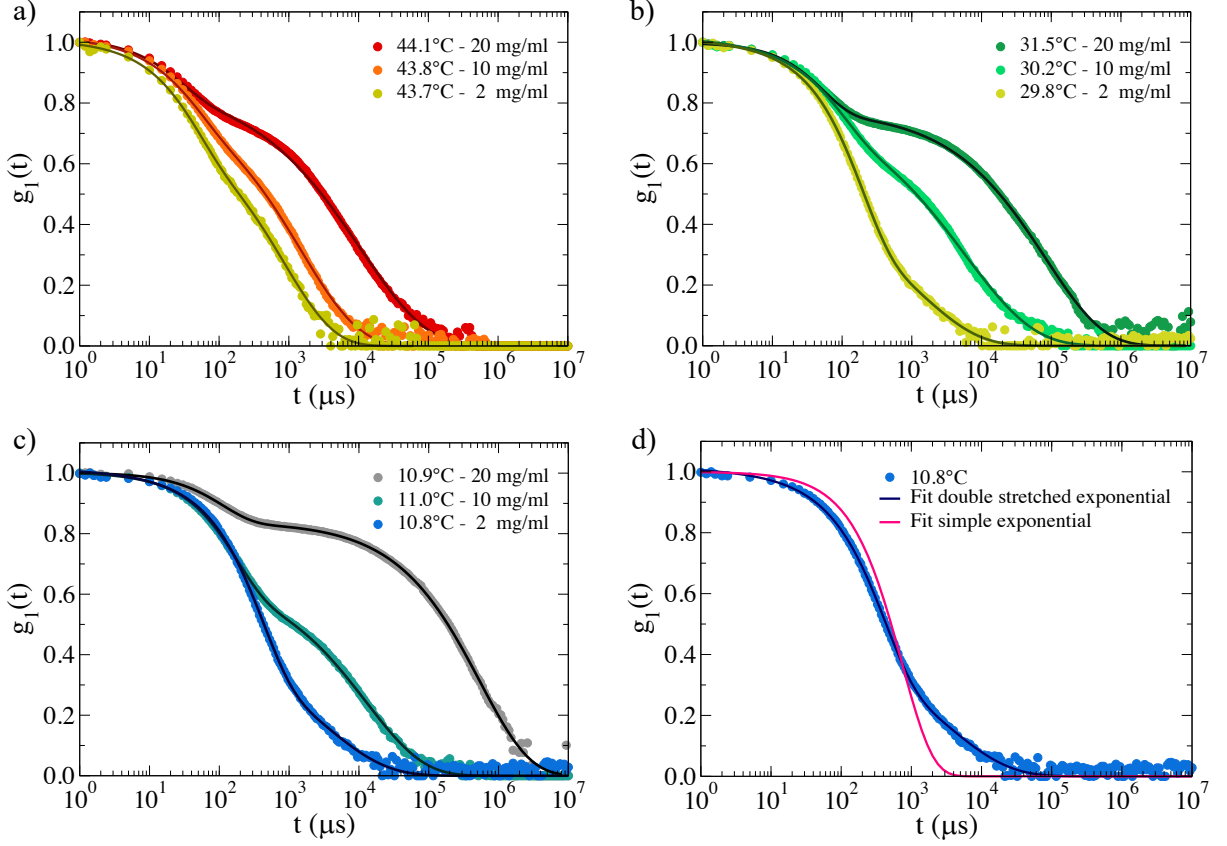

Figure S3: Panels a-c: Comparison of the experimental data (symbols) and the fit with Eq. (13) in the main document (full lines) for the autocorrelation functions $g_1(t)$ obtained for the three different investigated NS concentrations at high (a), intermediate (b), and low temperatures (c). Panel d: Comparison between the experimental data with both the simple exponential and double stretched exponential fits for the sample at $c = 2\,\mathrm{mg/ml}$, $T = 10.8°C$ (the experimental data and the double stretched exponential fit are the same as the ones reported in panel c).

F. Hydrodynamic radius from simulations

In Fig. S4, we show the hydrodynamic radius calculated from the simulations for $c = 2\,\mathrm{mg/ml}$ and $c = 20\,\mathrm{mg/ml}$ according to the method described in Ref. [6]. Briefly, for each cluster in the system, we computed the smallest convex set of points (the convex hull) that encloses the position of the bases of the NSs forming the clusters. Then, the gyration tensor is calculated using the center of mass of the triangles composing the convex hull. Finally, the gyration tensor is diagonalized, obtaining the three eigenvalues λ_1, λ_2, and λ_3, which are related to the three semi-axes $\{a\}$ of the ellipsoid as $a_i = \sqrt{3\lambda_i}$. The hydrodynamic radius is evaluated according to

$$R_h = \frac{2}{\int_0^\infty \left[(a_1^2 + \theta)(a_2^2 + \theta)(a_3^2 + \theta)\right]^{-\frac{1}{2}} \, \mathrm{d}\theta}. \tag{26}$$

As done for the radius of gyration in the main text, we calculated the ensemble average of the hydrodynamic radius over clusters of same size n and different simulation times.
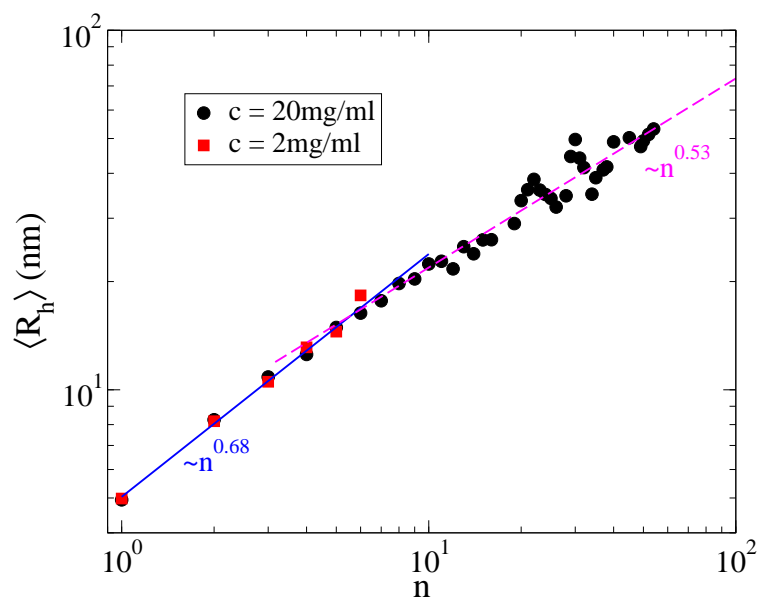

Figure S4: Hydrodynamic radii calculated from the simulations for $c = 2\,\mathrm{mg/ml}$ (black points) and $c = 20\,\mathrm{mg/ml}$ (red squares), respectively. Blue solid line and magenta dashed line are the power law fits to the data for $n \leq 5$ and $n > 5$, respectively.

[1] J. Zadeh, C. Steenberg, J. Bois, B. Wolfe, M. Pierce, A. Khan, R. Dirks and N. Pierce, *J. Comput. Chem.*, 2011, **32**, 170–173.
[2] P. Flory, *J. Am. Chem. Soc.*, 1941, **63**, 3083.
[3] M. Rubinstein and R. Colby, *Polymer Physics*, OUP Oxford, 2003.
[4] T. Hill, *Statistical Mechanics: Principles and Selected Applications*, Dover Publications, 1987.
[5] F. Sciortino, E. Bianchi, J. Douglas and P. Tartaglia, *J. Chem. Phys.*, 2007, **126**, 194903.
[6] L. Rovigatti, N. Gnan, A. Ninarello and E. Zaccarelli, *Macromolecules*, 2019, **52**, 4895.