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 μM (corresponding to the largest sample concentration $c = 20\,\mathrm{mg/ml}$) 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 ($T_{\mathrm{NS}} \simeq 77°\mathrm{C}$) and the formation of the interparticle bonds ($T_b \simeq 42°\mathrm{C}$).

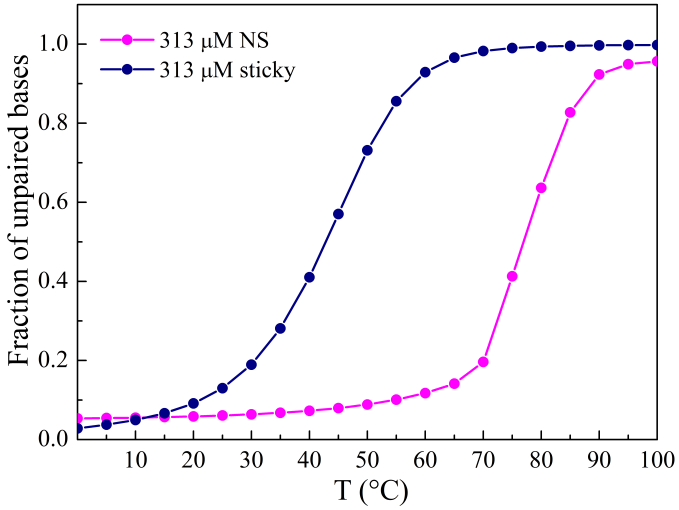

Figure S1: Melting curves calculated using NUPACK oligo analyzer [1] for both the NS arms (magenta) and the sticky tips (blue). Note that the self-assembly of the NS precedes on cooling the binding of the sticky sequences.

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 $3 \times 10^4\,\mathrm{K}$), 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 $c = 2\,\mathrm{mg/ml}$, $N = 1000$ (left) and $c = 20\,\mathrm{mg/ml}$, $N = 2000$ (right) when varying the number of bases used in the bond definition.

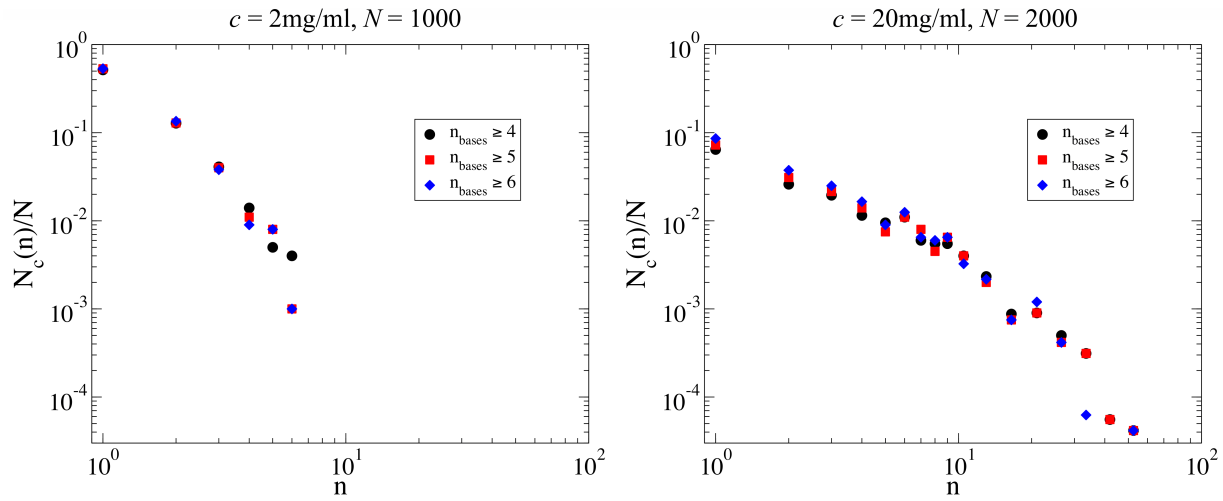

Figure S2: Effect of the choice of the minimum number of bases (n_{bases}) used to discriminate bonded and non-bonded nanostars.

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

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

so that

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

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

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

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),

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

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

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

such that

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

This time

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

so that

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

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

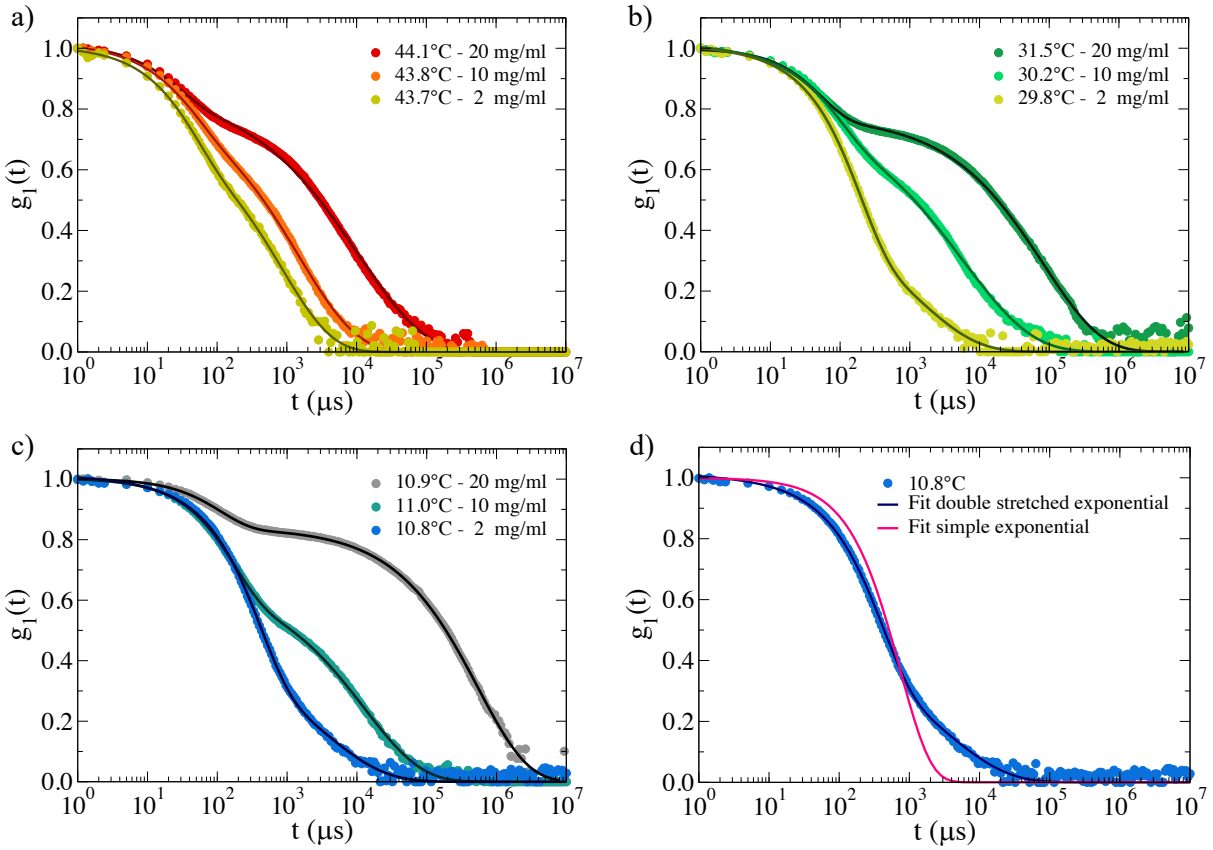

Figure S3: Panels a-c: Comparison of the experimental data (symbols) and the fit with Eq. (13) in the main document (full lines) for the autocorrelation functions $g_1(t)$ obtained for the three different investigated NS concentrations at high (a), intermediate (b), and low temperatures (c). Panel d: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

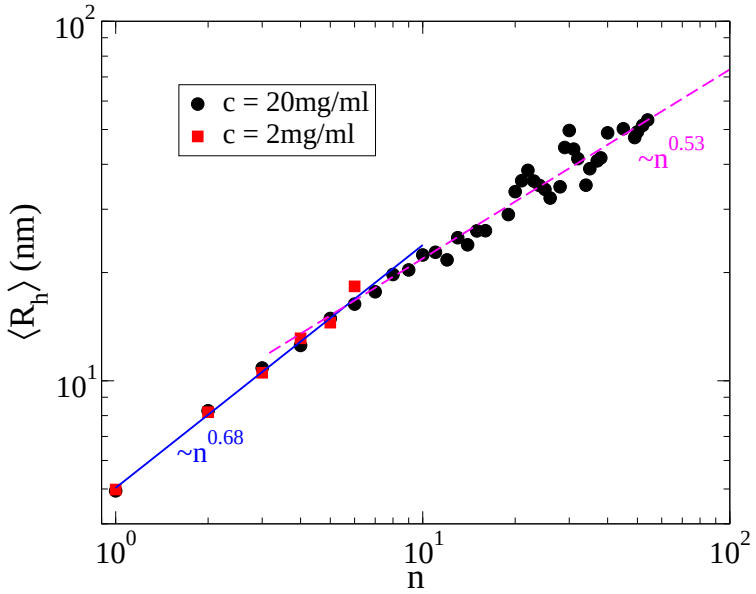

Figure S4: Hydrodynamic radii calculated from the simulations for $c = 2\,\mathrm{mg/ml}$ (black points) and $c = 20\,\mathrm{mg/ml}$ (red squares), respectively. Blue solid line and magenta dashed line are the power law fits to the data for $n \leq 5$ and $n > 5$, respectively.

[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.