ESI: Surface-triggered cascade reactions between DNA linkers direct single–staged colloidal deposition with a controllable number of layers

Pritam Kumar Jana* and Bortolo Matteo Mognetti*

Université Libre de Bruxelles (ULB), Interdisciplinary Center for Nonlinear Phenomena and Complex Systems, Campus Plaine, CP 231, Blvd. du Triomphe, B-1050 Brussels, Belgium

E-mail: Pritam.Kumar.Jana@ulb.ac.be; bmognett@ulb.ac.be

Supplementary Figures

Supplementary Video 1 Caption: Simulation trajectory showing the self-assembly of a single layer crystal using colloids with $N_{\mathrm{L}} = 40$ ligands (see Main Fig. 4).

Supplementary Video 2 Caption: Simulation trajectory showing the self-assembly of a two-layer crystal using colloids with $N_{\mathrm{L}} = 40$ ligands (see Main Fig. 4).

Supplementary Video 3 Caption: Simulation trajectory showing the self-assembly of a three-layer crystal using colloids with $N_{\mathrm{L}} = 40$ ligands (see Main Fig. 4).

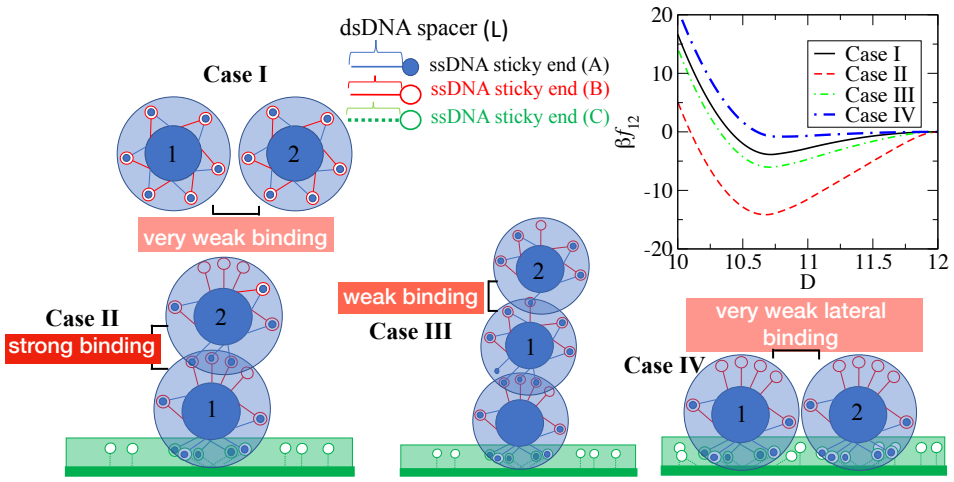

Figure S1: **Pair interaction of surface-bound particles (Case IV).** Two particles in direct contact with the surface express the same type of free linkers resulting in pair interactions weaker than in bulk. The system parameters are given in the caption of Main Fig. 1.

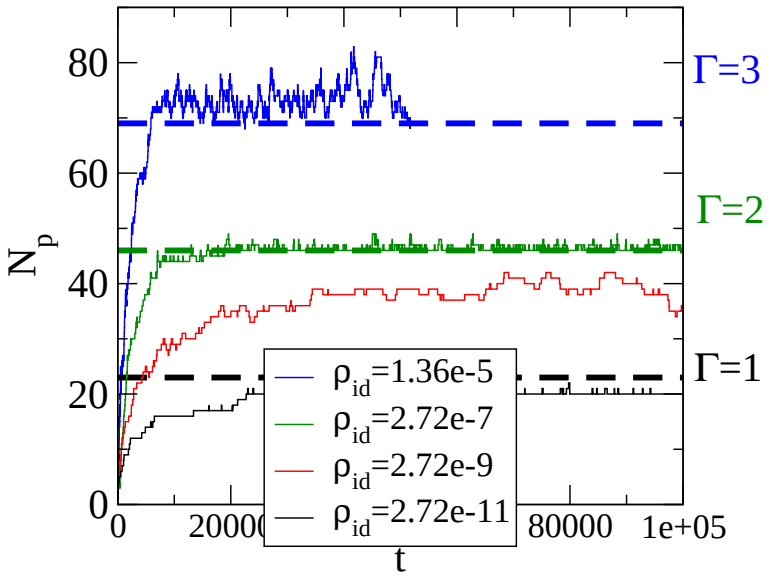

Figure S2: **Self-assembly directed by functionalized interfaces always leads to crystals with filled top layers.** We report the number of adsorbed colloids versus time obtained in simulations using different chemical potentials ($\rho_{\mathrm{id}} \sim \exp[\beta\mu]$). The dashed lines mark the number of colloids compatible with an integer number of layers, Γ. We used $N_{\mathrm{L}} = 40$, $\beta\Delta G_0 = \beta\Delta G_0^{\mathrm{s}} + 1 = -9$, and receptor density $\sigma_R = 1.8 \cdot L^{-2}$.

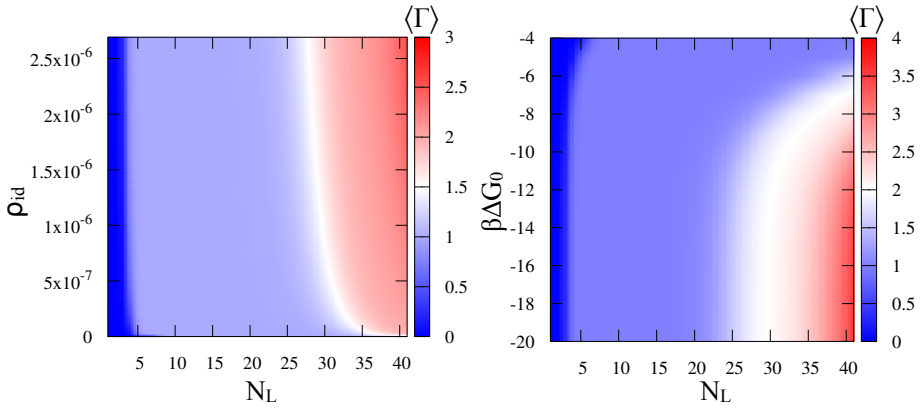

Figure S3: **Mean field theory predictions of the averaged number of self-assembled layers.** Effect of the number of ligands, the density of the gas phase, and ligand-ligand hybridization free-energy on the thickness of the self-assembled crystal. In panel (a) we used $\beta\Delta G_0 = -9$ while in panel (b) $\rho_{\rm id} = 2.72 \cdot 10^{-6} L^{-3}$. We used $\beta\Delta G_0^{\rm s} = \beta\Delta G_0 - 1$ and receptor density $\sigma_{\rm R} = 1.8 \cdot L^{-2}$.

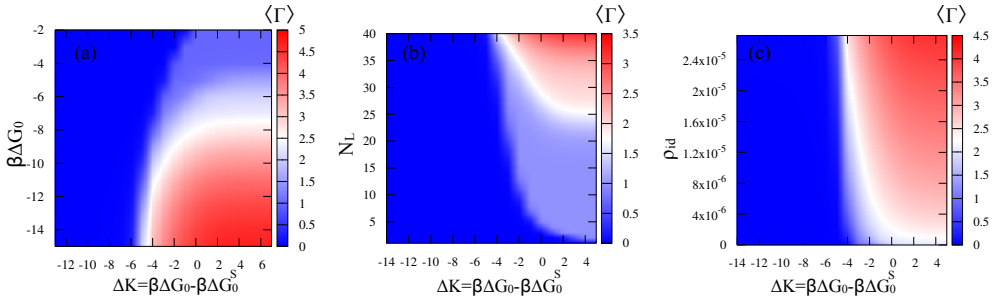

Figure S4: **Effect of the ligand–receptor hybridization free energy on the crystal thickness.** While decreasing $\Delta G_0^{\rm s}$, the averaged number of layers increases until reaching an asymptotic value. When not varied, we set $\beta\Delta G_0 = -9$, $N_{\rm L} = 40$, and $\rho_{\rm id} = 1.1 \cdot 10^{-5} \cdot L^{-3}$. For the receptor density we use $\sigma_{\rm R} = 1.8 \cdot L^{-2}$.

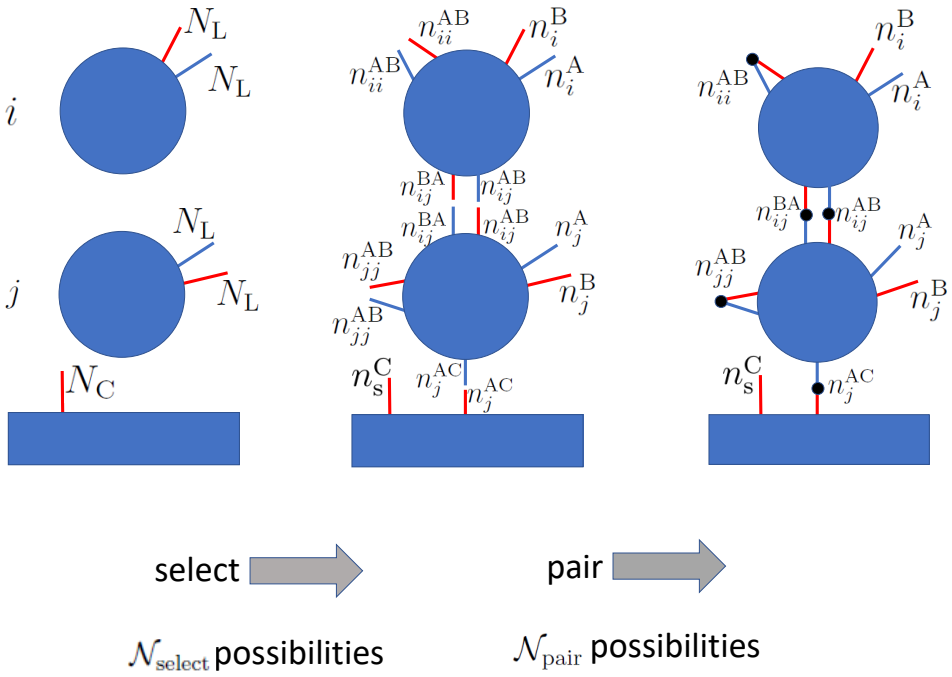

Figure S5: **Counting the number of ways of forming a given set of linkages $\{n\}$.** $\mathcal{N}_{\mathrm{select}}$ (see Main Eq. 6) is the number of ways of selecting the ligands/receptors to be used to form each type of linkage entering in $\{n\}$. $\mathcal{N}_{\mathrm{pair}}$ is the number of ways of binding the preselected ligands/receptors.

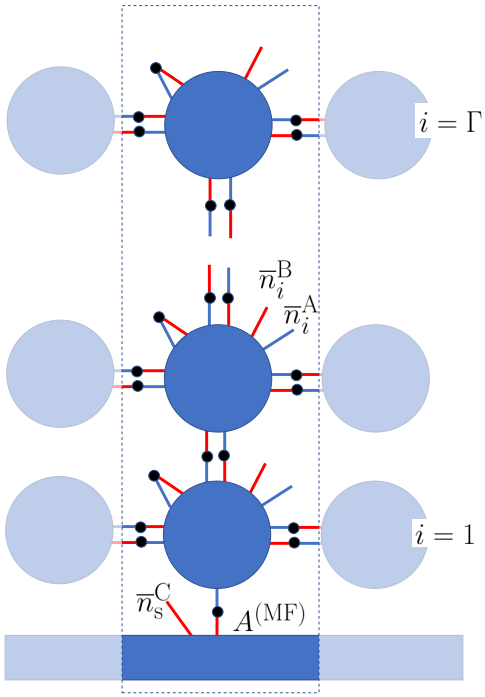

Figure S6: **Definitions of the variables used in the mean field calculation.** We consider fcc (111) crystals in which each colloid is surrounded by six particles belonging to the same layers (i) and three particles belonging to the upper ($i+1$) and/or lower ($i-1$) layer (for simplicity the figure reports at most four neighbors). Particles belonging to a given layer i feature the same type of linkages and carry the same number of free A and B binders, $\overline{n}_i^{\mathrm{A}}$ and $\overline{n}_i^{\mathrm{B}}$. The particle in the first layer faces a surface area $A^{(\mathrm{MF})}$ estimated using simulations. Accordingly, the total number of receptors $N_{\mathrm{C}}^{(\mathrm{MF})}$ is defined by the receptor density σ_{R}. In Main Figure 5 we used $A^{(\mathrm{MF})} = 10.31 \times 10.31 \cdot L^2$ while in Main Figure 6 $A^{(\mathrm{MF})} = 11.0 \times 11.0 \cdot L^2$. Particle–particle and particle–surface distances have been fixed to $11 \cdot L$ and $5.7 \cdot L$, respectively.

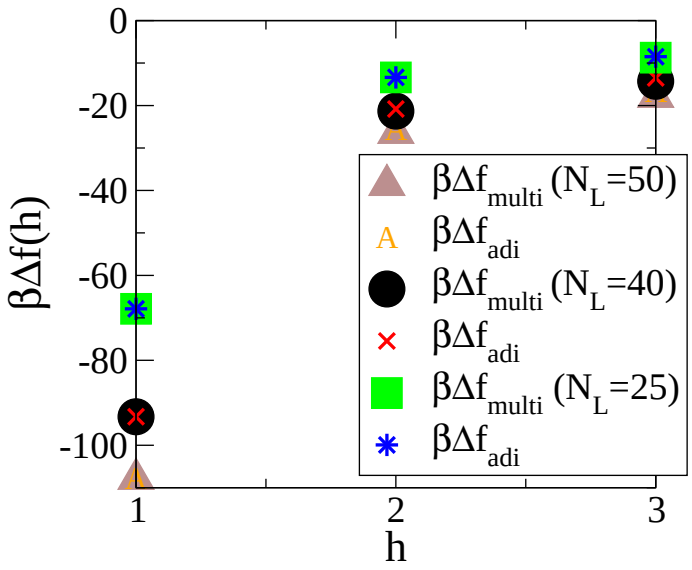

Figure S7: **Analytic approximations of the multivalent free energy match the results of the full theory.** For the three different number of ligands (N_{L}) considered in the present work, we compare the multivalent free energy as provided by the full theory ($\Delta f_{\mathrm{multi}}(h)$, see Eq. S25) with the analytic predictions of a simplified theory ($\Delta f_{\mathrm{adi}}(h)$, see Eq. S28) proving their consistency.

Supporting Information

S1 Calculation of the multivalent free energy of the system

We calculate the partition function of the system at a given colloids' configuration $\{\mathbf{r}\}$ and number of linkages $\{n\}$ (see Main Eq. 1) leading to the multivalent free energy $\mathcal{F}_{\rm multi}(\{\mathbf{r}\},\{n\}) = -k_BT\log\mathcal{Z}(\{\mathbf{r}\},\{n\})$. We start by computing the number of ways of making a certain set of linkages, $\{n\} = \{n_{ij}^{\rm AB}, n_{ji}^{\rm AB}, n_{ii}^{\rm AB}, n_i^{\rm AC}\}$ with $i = 1,\cdots N_{\rm p}$ and $i < j$. First, we count the number of ways, $\mathcal{N}_{\rm select}$, of selecting the binders (ligands and receptors) used to form a certain type of linkage (e.g., a bridge of type AB between particle i and particle j, see left and central panel in Fig. S5). The number of ways of selecting $n_{ij}^{\rm AB}$, $n_{ii}^{\rm AB}$, and $n_i^{\rm AC}$ A ligands on particle i to be used to form, respectively, the bridges with particle j ($j = 1,\cdots,N_{\rm p}$), the loops, and the surface–particle bridges is

$$\begin{aligned}\mathcal{N}^i_{\rm select,A} &= \binom{N_{\rm L}}{n_{ii}^{\rm AB}}\binom{N_{\rm L}-n_{ii}^{\rm AB}}{n_i^{\rm AC}}\prod_{j\neq i}\binom{N_{\rm L}-n_{ii}^{\rm AB}-n_i^{\rm AC}-\sum_{p=1,p\neq i}^{j-1}n_{ip}^{\rm AB}}{n_{ij}^{\rm AB}}\\ &= \frac{N_{\rm L}!}{n_i^{\rm A}!n_i^{\rm AC}!n_{ii}^{\rm AB}!\prod_{j\neq i}n_{ij}^{\rm AB}!}\,, \end{aligned} \tag{S1}$$

where $n_i^{\rm A}$ is the number of free A ligands

$$n_i^{\rm A} = N_{\rm L} - n_{ii}^{AB} - n_i^{\rm AC} - \sum_{j\neq i} n_{ij}^{\rm AB}\,. \tag{S2}$$

Similarly, the number of ways of selecting the B linkers on particle i and the receptors are

$$\mathcal{N}^i_{\rm select,B} = \frac{N_{\rm L}!}{n_i^{\rm B}!n_{ii}^{\rm AB}!\prod_{j\neq i}n_{ij}^{\rm BA}!} \qquad n_i^{\rm B} = N_{\rm L} - n_{ii}^{AB} - \sum_{j\neq i} n_{ij}^{\rm BA}, \tag{S3}$$

$$\mathcal{N}_{\rm select,C} = \frac{N_{\rm C}!}{n_{\rm s}^{\rm C}!\prod_{i=1}^{N_{\rm p}}n_i^{\rm AC}!} \qquad n_{\rm s}^{\rm C} = N_{\rm C} - \sum_i n_i^{\rm AC}, \tag{S4}$$

where $n_{\rm s}^{\rm C}$ is the number of free receptors on the surface. Finally, $\mathcal{N}_{\rm select}$ reads as

$$\mathcal{N}_{\rm select} = \mathcal{N}_{\rm select,C} \prod_{i=1}^{N_{\rm p}} \left[\mathcal{N}^i_{\rm select,A} \mathcal{N}^i_{\rm select,B} \right] . \tag{S5}$$

After selecting the ensemble of ligands/receptors to be used to form linkages $\{n\}$, we calculate the number of ways, $\mathcal{N}_{\rm pair}$, of reacting the pre–selected binders (see right panel of Fig. S5). Noticing that there are, for instance, $n_{ij}^{\rm AB}!$ ways of binding $n_{ij}^{\rm AB}$ A ligands on particle i with $n_{ij}^{\rm AB}$ B ligands on particle j, we obtain

$$\mathcal{N}_{\rm pair} = \prod_i \left[n_i^{\rm AC}! n_{ii}^{\rm AB}! \right] \prod_{i<j} \left[n_{ij}^{\rm AB}! n_{ij}^{\rm BA}! \right] . \tag{S6}$$

Finally, the combinatorial term associated to configurations with a given set of linkage $\{n\}$ is

$$\mathcal{N}_{\rm comb} = \mathcal{N}_{\rm select} \times \mathcal{N}_{\rm pair} = \frac{N_{\rm C}!}{n_{\rm s}^{\rm C}!} \prod_i \left[\frac{N_{\rm L}!}{n_i^{\rm A}!} \frac{N_{\rm L}!}{n_i^{\rm B}!} \frac{1}{n_{ii}^{\rm AB}!} \frac{1}{n_{\rm s}^{\rm C}!} \right] \prod_{i<j} \left[\frac{1}{n_{ij}^{\rm AB}!} \frac{1}{n_{ij}^{\rm BA}!} \right] \tag{S7}$$

To derive the final expression of $\mathcal{Z}$, we multiply $\mathcal{N}_{\rm comb}$ by the Boltzmann factor accounting for the hybridization free energy of binding a set $\{n\}$ of linkages (see Main text for the definitions)

$$\begin{aligned} \mathcal{Z} \;=\; & \left[\prod_{i=1}^{N_{\rm p}} \frac{N_{\rm L}! N_{\rm L}! e^{-\beta n_{ii}^{\rm AB} \Delta G_{ii}(\{\mathbf{r}\})}}{n_i^{\rm A}! n_i^{\rm B}! n_{ii}^{\rm AB}!} \right] \left[\prod_{j<q} \frac{e^{-\beta (n_{jq}^{\rm AB} + n_{jq}^{\rm BA}) \Delta G_{jq}(\{\mathbf{r}\})}}{n_{jq}^{\rm AB}! n_{jq}^{\rm BA}!} \right] \left[\frac{N_{\rm C}!}{n_{\rm s}^{\rm C}!} \prod_{i=1}^{N_{\rm p}} \frac{e^{-\beta n_i^{\rm AC} \Delta G_i^s(\{\mathbf{r}\})}}{n_i^{\rm AC}!} \right] \\ & \cdot \mathcal{Z}_{T=\infty}(\{\mathbf{r}\}) . \end{aligned} \tag{S8}$$

$\mathcal{Z}_{T=\infty}$ denotes the partition function of the system when no linkages are formed (as found at high temperature). $\mathcal{Z}_{T=\infty}$ is related to entropic, repulsive forces due to the reduction of the configurational space of the DNA linkers compressed by two approaching surfaces. Neglecting excluded volume interactions between linkers, as often done when modeling DNA mediated interactions, and defining by Ω_0 and $\Omega_i(\{\mathbf{r}\})$ the volume of the configurational

space available, respectively, to a single ligand tethered to particle i at infinite dilution and at finite density (similar definitions follow for the configurational space of receptors tethered to the surface, $\Omega_0^{\rm s}$ and $\Omega^{\rm s}(\{\mathbf{r}\})$) we find

$$Z_{T=\infty}(\{\mathbf{r}\}) \;=\; e^{-\beta F_{T=\infty}(\{\mathbf{r}\})} = \left(\frac{\Omega^{\rm s}(\{\mathbf{r}\})}{\Omega_0^{\rm s}}\right)^{N_{\rm C}} \cdot \prod_{i=1}^{N_{\rm P}} \left(\frac{\Omega_i(\{\mathbf{r}\})}{\Omega_0}\right)^{N_{\rm L}+N_{\rm L}} e^{-\beta V(\{\mathbf{r}\})}\,. \quad \text{(S9)}$$

In the previous expression, $V(\{\mathbf{r}\})$ accounts for additional (e.g. electrostatic) interactions between colloids. We use $V(\{\mathbf{r}\}))$ to regularize hard–core repulsions with negligible effects on the final results. Below (see Sec. S1.1), we report the expression of $V(\{\mathbf{r}\})$ that has been used in this work.

In this work we consider reactive sequences tethered to particles' surfaces through short, thin rods of double-stranded DNA of length L. When L is much smaller than the radius of the particles, the reactive sequences of unbound linkers are uniformly distributed within the layer of thickness L surrounding the tethering surfaces. We then have $\Omega_0 = 4\pi R^2 L$, $\Omega_0^{\rm s} = \mathcal{A}L$ (where $\mathcal{A}$ is the area of the surface). Similarly, the configurational volumes defining the hybridization free energies (see Main Eqs. 4) read as

$$\begin{aligned}
\Omega_i(\{\mathbf{r}\}) &= \Omega_0 - e_i^{\rm s}(r_{i,z}) - \sum_{j\in v(i)} e_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) \\
\Omega^{\rm s}(\{\mathbf{r}\}) &= \Omega_0^{\rm s} - \sum_{i\in v_{\rm s}} k_i^{\rm s}(r_{i,z})
\end{aligned} \quad \text{(S10)}$$

where e_{ij} and $e_i^{\rm s}$ are, respectively, the volume excluded to the reactive sequence of a linker tethered to particle i by the presence of particle j and the surface (see Main Fig. 2). $k_i^{\rm s}$ is the volume excluded to a reactive sequence tethered to the surface by the presence of particle i (see Main Fig. 2). $v(i)$ and $v_{\rm s}$ are the lists of particles in direct contact with ligands on particle i and receptors on the surface. Similarly, the configurational space of bound sequences (Ω_{ij} and $\Omega_i^{\rm s}$) is the volume of the overlapping regions spanned by the reacting sequences before binding (see Main Fig. 2). We report the explicit expression of the terms

appearing in Main Eqs. 4 and Eqs. S10 in Sec. S1.2.

At given colloid positions, $\{\mathbf{r}\}$, the most likely number of linkages featured by the system, $\{\overline{n}\}$, are calculated by maximizing the multivalent free energy

$$\frac{\partial}{\partial\{n\}}\mathcal{F}_{\rm multi}(\{n\})|_{\{n\}=\{\overline{n}\}} = 0\,. \tag{S11}$$

The previous set of equations, along with the definition of $\mathcal{F}_{\rm multi}$, Eq. S8, lead to the chemical equilibrium equations reported in Main Eqs. 3. Main Eqs. 3 become equivalent to the following set of equations for the number of unbound linkers

$$\begin{aligned}
\overline{n}_i^{\rm A} &= \frac{N_{\rm L}}{1+\overline{n}_{\rm s}^{\rm C}e^{-\beta\Delta G_i^{\rm s}(\{\mathbf{r}\})}+\sum_{j\in v(i)}\overline{n}_j^{\rm B}e^{-\beta\Delta G_{ij}(\{\mathbf{r}\})}}\\
\overline{n}_i^{\rm B} &= \frac{N_{\rm L}}{1+\sum_{j\in v(i)}\overline{n}_j^{\rm A}e^{-\beta\Delta G_{ij}(\{\mathbf{r}\})}}\\
\overline{n}_{\rm s}^{\rm C} &= \frac{N_{\rm C}}{1+\sum_{j\in v_{\rm s}}\overline{n}_j^{\rm A}e^{-\beta\Delta G_j^{\rm s}(\{\mathbf{r}\})}}\,,
\end{aligned} \tag{S12}$$

that are used to implement self-consistent calculations in our simulations (see Sec. S3). When written in term of the stationary number of linkages, the multivalent free energy simplifies into the expression reported in Main Eq. 5.

S1.1 Modeling hard–core repulsion

We use smooth pair potentials to regularize particle–particle and particle–surface hard–core repulsions ($V(\{\mathbf{r}\})$ in the Methods section of the main text). Following previous investigations, the repulsion between colloids is modeled using

$$V_{\rm pp}(r_{ij}) = \begin{cases} 500\log\left(1-\frac{e_{ij}(r_{ij},0.75\cdot L)}{4\pi R^2\cdot(0.75\cdot L)}\right) & r_{ij} < 2\cdot R+0.75\cdot L\\ 0 & r_{ij}\geq 2\cdot R+0.75\cdot L \end{cases} \tag{S13}$$

where r_{ij} is the distance between particle i and j, and e_{ij} is defined below (Eq. S18). Similarly, if $r_{i,\mathrm{z}}$ is the distance of particle i from the surface, the surface-particle repulsion is modeled using

$$V_{\mathrm{ps}}(r_{i,\mathrm{z}}) = \begin{cases} 500 \log\left(1 - \frac{e_i^{\mathrm{s}}(r_{i,\mathrm{z}}, 0.75 \cdot L)}{4\pi R^2 \cdot (0.75 \cdot L)}\right) & r_{i,\mathrm{z}} < R + 0.75 \cdot L \\ 0 & r_{i,\mathrm{z}} \geq R + 0.75 \cdot L \end{cases}, \tag{S14}$$

where e_i^{s} is defined in Eq. S19. Finally

$$V(\{\mathbf{r}\}) = \sum_{i<j} V_{\mathrm{pp}}(r_{ij}) + \sum_i V_{\mathrm{ps}}(r_{i,\mathrm{z}})\,. \tag{S15}$$

Such smooth regularizations allow using larger integration steps Δt (see the Methods section of the main text) with negligible effects on the results of the manuscript. The latter claim follows from the fact that the typical surface–to–surface distance is comparable with L while $V_{\mathrm{pp}}(r_{ij}) > 0$ and $V_{\mathrm{ps}}(r_{i\mathrm{z}}) > 0$ only when $r < 0.75 \cdot L$. V_{pp} and V_{ps} are effective potentials resulting from coating particles with inert strands (not carrying sticky ends) of length $0.75 \cdot L$. This observation justifies the particular choice of V given the fact that, in experiments, inert binders are often used to screen non-selective attractions (e.g., van der Waals forces).

S1.2 Configurational terms

Below we report the analytic expressions of the overlapping volumes defined in Main Fig. 2*b* and used to calculate the hybridization free energies and the *on* rates (see the Methods

section in the main text)

$$\Omega_{ij}(r_{ij}, L) = v_{\mathrm{par,par}}(r_{ij}, R+L, R+L) - 2v_{\mathrm{par,par}}(r_{ij}, R, R) \quad \text{(S16)}$$

$$\Omega_i^{\mathrm{s}}(r_{i,\mathrm{z}}, L) = v_{\mathrm{par,surf}}(r_{i,\mathrm{z}} - L, R+L) - v_{\mathrm{par,surf}}(r_{i,\mathrm{z}}, R+L) - v_{\mathrm{par,surf}}(r_{i,\mathrm{z}} - L, R) \quad \text{(S17)}$$

$$e_{ij}(r_{ij}, L) = v_{\mathrm{par,par}}(r_{ij}, R+L, R) \quad \text{(S18)}$$

$$e_i^{\mathrm{s}}(r_{i,\mathrm{z}}, L) = v_{\mathrm{par,surf}}(r_{i,\mathrm{z}}, R+L) \quad \text{(S19)}$$

$$k_i^{\mathrm{s}}(r_{i,\mathrm{z}}, L) = v_{\mathrm{par,surf}}(r_{i,\mathrm{z}} - L, R) \quad \text{(S20)}$$

where $v_{\mathrm{par,par}}(r, R_1, R_2)$ and $v_{\mathrm{par,surf}}(r, R)$ are, respectively, the overlapping volume between two spheres of radius R_1 and R_2 placed at distance r and the volume of a spherical cap of radius R with base placed at distance r from the center of the sphere:

$$v_{\mathrm{par,par}}(r, R_1, R_2) = \frac{\pi}{12r}\left(R_1 + R_2 - r\right)^2\left(r^2 + 2rR_1 + 2rR_2 - 3R_1^2 - 3R_2^2 + 6R_1R_2\right) \quad \text{(S21)}$$

$$v_{\mathrm{par,surf}}(r, R) = \frac{\pi}{3}\left(R - r\right)^2\left(2R + r\right) \quad \text{(S22)}$$

S2 Mean Field Theory

We detail the calculation of f_{multi} (see Main Eqs. 9) used to predict the probability of self-assembling crystals with Γ layers (see $P(\Gamma)$ defined in Main Eq. 10). $f_{\mathrm{multi}}(\Gamma)$ is the free energy of an fcc (111) crystallite comprising Γ layers as compared to a reference state in which particles are isolated and only feature loops, normalized by the number of particles in direct contact with the functionalized surface. We first calculate the number of free binders and free receptors on the particles belonging to layer i along with the number of free receptors ($\overline{n}_i^{\mathrm{A}}$, $\overline{n}_i^{\mathrm{B}}$, and $\overline{n}_{\mathrm{s}}^{\mathrm{C}}$ in Fig. S6). We consider particles distributed on an fcc (111) crystal with fixed particle–particle and particle–surface distance. Therefore, the free energy of making inter–particle bridges (ΔG^{b}) is constant and is calculated using Main Eq. 4. Similarly, ΔG^{l} and $\Delta G^{\mathrm{b,s}}$ are the hybridization free energies of forming loops and particle–surface bridges.

We calculate $\Delta G^{\mathrm{b,s}}$ using a surface area $A^{(\mathrm{MF})}$ corresponding to the averaged surface *per* surface–bound particle as sampled in a representative simulation (notice that $A^{(\mathrm{MF})}$ affects $\Delta G^{\mathrm{b,s}}$ through Ω^{s}, see Main Eq. 4). Accordingly, we fix the number of receptors $N_{\mathrm{C}}^{(\mathrm{MF})}$ to $N_{\mathrm{C}}^{(\mathrm{MF})} = \sigma_R \cdot A^{(\mathrm{MF})}$, where σ_R is the receptor density.

Using Eqs. S12 we write the total number of free binders on the surface and on the particles belonging to the first layer as

$$\begin{aligned}
\overline{n}_{\mathrm{s}}^{\mathrm{C}} &= \frac{N_{\mathrm{C}}^{(\mathrm{MF})}}{1+\overline{n}_i^{\mathrm{A}}\exp[-\beta\Delta G^{\mathrm{b,s}}]} \\
\overline{n}_1^{\mathrm{A}} &= \frac{N_{\mathrm{L}}}{1+\overline{n}_1^{\mathrm{B}}(6\cdot\exp[-\beta\Delta G^{\mathrm{b}}]+\exp[-\beta\Delta G^{\mathrm{l}}])+3\cdot\overline{n}_2^{\mathrm{B}}\exp[-\beta\Delta G^{\mathrm{b}}]+\overline{n}_{\mathrm{s}}^{\mathrm{C}}\exp[-\beta\Delta G^{\mathrm{b,s}}]} \\
\overline{n}_1^{\mathrm{B}} &= \frac{N_{\mathrm{L}}}{1+\overline{n}_1^{\mathrm{A}}(6\cdot\exp[-\beta\Delta G^{\mathrm{b}}]+\exp[-\beta\Delta G^{\mathrm{l}}])+3\cdot\overline{n}_2^{\mathrm{A}}\exp[-\beta\Delta G^{\mathrm{b}}]}
\end{aligned}$$

where in the expression of $\overline{n}_1^{\mathrm{A}}$ and $\overline{n}_1^{\mathrm{B}}$ we set $\overline{n}_2^{\mathrm{A}} = \overline{n}_2^{\mathrm{B}} = 0$ when calculating the free energy of single–layer crystals ($\Gamma = 1$ in Fig. S6). For particles in the intermediate layers $1 < i < \Gamma$ we have

$$\begin{aligned}
\overline{n}_i^{\mathrm{A}} &= \frac{N_{\mathrm{L}}}{1+\overline{n}_i^{\mathrm{B}}(6\cdot\exp[-\beta\Delta G^{\mathrm{b}}]+\exp[-\beta\Delta G^{\mathrm{l}}])+3\cdot(\overline{n}_{i-1}^{\mathrm{B}}+\overline{n}_{i+1}^{\mathrm{B}})\exp[-\beta\Delta G^{\mathrm{b}}]} \\
\overline{n}_i^{\mathrm{B}} &= \frac{N_{\mathrm{L}}}{1+\overline{n}_i^{\mathrm{A}}(6\cdot\exp[-\beta\Delta G^{\mathrm{b}}]+\exp[-\beta\Delta G^{\mathrm{l}}])+3\cdot(\overline{n}_{i-1}^{\mathrm{A}}+\overline{n}_{i+1}^{\mathrm{A}})\exp[-\beta\Delta G^{\mathrm{b}}]}
\end{aligned}$$

while for the last layer (if $\Gamma > 1$)

$$\begin{aligned}
\overline{n}_\Gamma^{\mathrm{A}} &= \frac{N_{\mathrm{L}}}{1+\overline{n}_\Gamma^{\mathrm{B}}(6\cdot\exp[-\beta\Delta G^{\mathrm{b}}]+\exp[-\beta\Delta G^{\mathrm{l}}])+3\cdot\overline{n}_{\Gamma-1}^{\mathrm{B}}\exp[-\beta\Delta G^{\mathrm{b}}]} \\
\overline{n}_\Gamma^{\mathrm{B}} &= \frac{N_{\mathrm{L}}}{1+\overline{n}_\Gamma^{\mathrm{A}}(6\cdot\exp[-\beta\Delta G^{\mathrm{b}}]+\exp[-\beta\Delta G^{\mathrm{l}}])+3\cdot\overline{n}_{\Gamma-1}^{\mathrm{A}}\exp[-\beta\Delta G^{\mathrm{b}}]}
\end{aligned}$$

We define by $\overline{n}_0^{\mathrm{A}}$ and $\overline{n}_0^{\mathrm{B}}$ ($\overline{n}_0^{\mathrm{A}} = \overline{n}_0^{\mathrm{B}}$) the number of free binders of particles isolated in bulk (therefore featuring $N_{\mathrm{L}} - \overline{n}_0^{\mathrm{A}}$ loops). $\overline{n}_0^{\mathrm{A}}$ or $\overline{n}_0^{\mathrm{B}}$ is calculated by setting $\Delta G^{\mathrm{b}} = +\infty$ and $\Delta G^{\mathrm{b,s}} = +\infty$ in one of the previous equations.

The free energy of Γ isolated particles in bulk reads as (see Main Eq. 5)

$$\beta F_0(\Gamma) = \Gamma \left[N_L \log \frac{\overline{n}_0^{\rm A}}{N_{\rm L}} + N_L \log \frac{\overline{n}_0^{\rm B}}{N_{\rm L}} + (N_{\rm L} - \overline{n}_0^{\rm A}) \right] , \tag{S23}$$

where $N_{\rm L} - \overline{n}_0^{\rm A}$ is the number of loops featured by each particle. On the other hand, the free energy *per* surface–bound particle of crystals with Γ layers (see Fig. S6) is given by (see Main Eq. 1)

$$\beta F(\Gamma) \;\; = \;\; N_{\rm C} \log \frac{\overline{n}_{\rm s}^{\rm C}}{N_{\rm C}} + \frac{N_{\rm C} - \overline{n}_{\rm s}^{\rm C}}{2} + \sum_{i=1}^{\Gamma} \left[N_{\rm L} \log \frac{\overline{n}_i^{\rm A}}{N_{\rm L}} + N_{\rm L} \log \frac{\overline{n}_i^{\rm B}}{N_{\rm L}} + \frac{(N_{\rm A} + N_{\rm B}) - \overline{n}_i^{\rm A} - \overline{n}_i^{\rm B}}{2} \right] . \tag{S24}$$

Notice that in the configuration of Fig. S6 particles are sufficiently distanced that $F_{T=\infty} = 0$. Finally, $f_{\rm multi}(\Gamma)$ (Main Eq. 9) used in the definition of P (Main Eq. 10) and $\Delta f_{\rm multi}(\Gamma)$ (see Main Fig. 5a-c) are given by

$$f_{\rm multi}(\Gamma) = F(\Gamma) - F_0(\Gamma) \qquad\qquad \Delta f_{\rm multi}(\Gamma) = f_{\rm multi}(\Gamma) - f_{\rm multi}(\Gamma - 1) \, . \tag{S25}$$

To calculate the equilibrium layer distribution $P(\Gamma)$, we also need to estimate the entropic loss of caging colloids from the gas phase into a site of the crystal. We employ a cell model in which we assign a configurational space volume equal to v_0 to each particle in the solid phase. Following Main Refs. [42,43] we use $v_0 = (L/2)^3$, where we identify L with the interaction range of a square well potential. We study the sensitivity of our results to v_0 in Main Fig. 8. The probability of assembling Γ layers from a diluted colloidal suspension at density $\rho_{\rm id}$ reads as $P(\Gamma) = 1/Z_{\rm MFT} \cdot \exp[-\beta f_{\rm multi}(\Gamma)](\rho_{\rm id} v_0)^{\Gamma}$, where $\rho_{\rm id} \sim \exp[\beta\mu]$, μ is the chemical potential of the particles, and $\rho_{\rm id}$ is small enough to justify an ideal representation of the gas phase. $Z_{\rm MFT}$ is a normalization factor that is well defined (i.e. $Z_{\rm MFT} = \sum_{\Gamma=0}^{\infty} \exp[-\beta f_{\rm multi}(\Gamma)](\rho_{\rm id} v_0)^{\Gamma} < \infty$) conditional on the gas phase being stable in bulk. To extract the phase boundary in bulk, we notice that for $\Gamma \to \infty$ surface effects are negligible and $f_{\rm multi}(\Gamma)$ reads as $\Delta f(\infty) \cdot \Gamma$,

where $\Delta f_{\rm multi}(\infty)$, $\Delta f_{\rm multi}(\infty) = \lim_{\Gamma\to\infty} f_{\rm multi}(\Gamma) - f_{\rm multi}(\Gamma - 1)$, is the multivalent free energy *per* particle in an fcc crystal as compared to the gas phase. The gas–solid boundary in bulk (see Main Figs. 3*b*, 3*c*, and 4) is given by the relation $\exp[-\beta\Delta f_{\rm multi}(\infty)]\rho_{\rm id}v_0 = 1$ or $\beta\Delta f_{\rm multi}(\infty) = \log(\rho_{\rm id}v_0)$ that, in the diluted limit $\rho_{\rm id} \to 0$, matches existing cell models Main Ref. [43].

S2.1 Analytic predictions

In this section we extract compact analytic expressions allowing to estimate $\Delta f_{\rm multi}$ (see Eq. S25). We consider the low temperature regime in which $\Delta G_0 \to -\infty$. Using Main Eq. 5, we calculate the statistical weight of, respectively, particle–surface and particle–particle bridges divided by the statistical weight of intra–particle loops as

$$\begin{aligned} \chi^{\rm b,s} &= \frac{\exp[-\beta\Delta G^{\rm b,s}]}{\exp[-\beta\Delta G^{\rm l}]} = \frac{\Omega_i^{\rm s}}{\Omega_0^{\rm s}} \\ \chi^{\rm b} &= \frac{\exp[-\beta\Delta G^{\rm b}]}{\exp[-\beta\Delta G^{\rm l}]} = \frac{\Omega_{ij}}{\Omega_0} \ . \end{aligned} \tag{S26}$$

In the previous expressions, we considered that for the colloidal arrangement used in the MFT (see Fig. S6) the configurational space of free binders is not excluded by any colloid or the surface ($\Omega_i = \Omega_0$ and $\Omega^{\rm s} = \Omega_0^{\rm s}$). Note that Ω_{ij} and $\Omega_i^{\rm s}$ are not a function of the specific particle i given that particle–particle and particle–surface distances are kept constant (see Eq. S16). Once the first layer of particles is formed, each particle will present a number of free ligands equal to

$$M_1 = \frac{\sqrt{(\chi^{\rm b,s})^2(N_{\rm C}^{\rm (MF)} - N_{\rm L})^2 + 4\chi^{\rm b,s}N_{\rm C}^{\rm (MF)}N_{\rm L}(1+6\chi^{\rm b})} - \chi^{\rm b,s}(N_{\rm C}^{\rm (MF)} + N_{\rm L})}{2(1 - \chi^{\rm b,s} + 6\chi^{\rm b})} \ . \tag{S27}$$

We now assume that when particles are added to the second layer, the number of particle–surface bridges as well as the number of lateral bridges between particles in the first layer do

not change. Such approximation allows calculating $\Delta f_{\rm multi}(2)$ only using M_1. Similarly, by calculating the number of free linkers featured by particles in the second layer (M_2, which is only function of M_1), we can re-iterate the calculation of $\Delta f_{\rm multi}(3)$ for particles in the third layer. In general, the free-energy gain of adding layer i when particles in layer $i-1$ express M_{i-1} free linkers is

$$\begin{aligned}\Delta f_{\rm adi}(i) &= \frac{3\chi^{\rm b}\left[M_{i-1}-3(N_{\rm L}+4\chi^{\rm b}N_{\rm L})\right]-\Lambda}{4(1+3\chi^{\rm b})(1+6\chi^{\rm b})}+N_{\rm L}\log\frac{1}{3\chi^{\rm b}(M_{i-1}-N_{\rm L})+\Lambda}\\ &\quad +N_{\rm L}\log\frac{\Lambda-3\chi^{\rm b}(M_{i-1}+N_{\rm L})}{1+3\chi^{\rm b}}\end{aligned} \tag{S28}$$

where

$$\Lambda = \sqrt{3\chi^{\rm b}}\sqrt{4M_{i-1}N_{\rm L}+3\chi^{\rm b}\left[(M_{i-1})^2+6M_{i-1}N_{\rm L}+N_{\rm L}^2\right]}\,. \tag{S29}$$

The number of free ligands *per* particle expressed by layer i before attaching layer $i+1$ reads as

$$M_i = \frac{\Lambda-3\chi^{\rm b}M_i-3\chi^{\rm b}N_L}{2(1+3\chi^{\rm b})}\,. \tag{S30}$$

Fig. S7 shows that the results of the analytic theory match the MFT predictions.

S3 Simulation details

We first calculate the force acting on particle i, $\mathbf{f}_i$ in Main Eq. 6 and 7. When using the implicit (IMP) scheme (see Main Sec. 2.2), the numbers of possible linkages are fixed to their most likely values, $\{n\}=\{\overline{n}\}$ (see Main Eqs. 12, 13). The force acting on particle i then

reads as

$$\begin{aligned}
\beta\mathbf{f}_i &= -\beta\nabla_{\mathbf{r}_i}\mathcal{F}_{\rm multi}(\{\overline{n}\},\{\mathbf{r}\}) = -\nabla_{\mathbf{r}_i}\log\mathcal{Z}(\{\overline{n}\},\{\mathbf{r}\}) \\
&= -\frac{\partial}{\partial\{n\}}\beta\mathcal{F}_{\rm multi}(\{n\},\{\mathbf{r}\})|_{\{n\}=\{\overline{n}\}}\nabla_{\mathbf{r}_i}\{\overline{n}\} - \frac{\partial}{\partial\{\Delta G\}}\beta\mathcal{F}_{\rm multi}(\{n\},\{\mathbf{r}\})\nabla_{\mathbf{r}_i}\{\Delta G\}|_{\{n\}=\{\overline{n}\}} \\
&\quad -\beta\frac{\partial}{\partial\mathbf{r}_i}\mathcal{F}_{\rm multi}(\{n\},\{\mathbf{r}\})|_{\{n\}=\{\overline{n}\}} \\
&= -\frac{\partial}{\partial\{\Delta G\}}\beta\mathcal{F}_{\rm multi}(\{n\},\{\mathbf{r}\})\nabla_{\mathbf{r}_i}\{\Delta G\}|_{\{n\}=\{\overline{n}\}} - \beta\frac{\partial}{\partial\mathbf{r}_i}F_{T=\infty}(\{\mathbf{r}\}) \qquad \text{(S31)}
\end{aligned}$$

where the second equality follows from the saddle point equations (Main Eq. 2) and the fact that the only direct dependency of $\mathcal{F}_{\rm multi}$ on $\{\mathbf{r}\}$ is due to $F_{T=\infty}$ (see Eqs. S8 and S9). In particular, we find

$$\begin{aligned}
\beta\mathbf{f}_i &= -\sum_{j=1}^{N_{\rm p}}\left[\overline{n}_{jj}^{\rm AB}\nabla_{\mathbf{r}_i}\beta\Delta G_{jj}(\{\mathbf{r}\}) + \overline{n}_{j}^{\rm AC}\nabla_{\mathbf{r}_i}\beta\Delta G_{j}^{s}(\{\mathbf{r}\}) - (N_{\rm L}+N_{\rm L})\frac{\nabla_{\mathbf{r}_i}\Omega_j(\{\mathbf{r}\})}{\Omega_j(\{\mathbf{r}\})}\right] \\
&\quad -\sum_{1\le j<q\le N_{\rm p}}(\overline{n}_{jq}^{\rm AB}+\overline{n}_{jq}^{\rm BA})\nabla_{\mathbf{r}_i}\beta\Delta G_{jq}(\{\mathbf{r}\}) + N_{\rm C}\frac{\nabla_{\mathbf{r}_i}\Omega^{\rm s}(\{\mathbf{r}\})}{\Omega^{\rm s}(\{\mathbf{r}\})} - \nabla_{\mathbf{r}_i}\beta V(\{\mathbf{r}\})\,. \qquad \text{(S32)}
\end{aligned}$$

By using the definitions of ΔG we find

$$\beta\nabla_{\mathbf{r}_i}\Delta G_{jj}(\{\mathbf{r}\}) = \frac{\nabla_{\mathbf{r}_i}\Omega_j(\{\mathbf{r}\})}{\Omega_j(\{\mathbf{r}\})} \qquad \text{(S33)}$$

$$\beta\nabla_{\mathbf{r}_i}\Delta G_{jq}(\{\mathbf{r}\}) = \frac{\nabla_{\mathbf{r}_i}\Omega_j(\{\mathbf{r}\})}{\Omega_j(\{\mathbf{r}\})} + \frac{\nabla_{\mathbf{r}_i}\Omega_q(\{\mathbf{r}\})}{\Omega_q(\{\mathbf{r}\})} - \frac{\nabla_{\mathbf{r}_i}\Omega_{jq}(\{\mathbf{r}\})}{\Omega_{jq}(\{\mathbf{r}\})} \qquad \text{(S34)}$$

$$\beta\nabla_{\mathbf{r}_i}\Delta G_{j}^{\rm s}(\{\mathbf{r}\}) = \frac{\nabla_{\mathbf{r}_i}\Omega_j(\{\mathbf{r}\})}{\Omega_j(\{\mathbf{r}\})} + \frac{\nabla_{\mathbf{r}_i}\Omega^{\rm s}(\{\mathbf{r}\})}{\Omega^{\rm s}(\{\mathbf{r}\})} - \frac{\nabla_{\mathbf{r}_i}\Omega_j^{\rm s}(\{\mathbf{r}\})}{\Omega_j^{\rm s}(\{\mathbf{r}\})} \qquad \text{(S35)}$$

$$\beta\nabla_{\mathbf{r}_i}\Delta G_{\rm s}(\{\mathbf{r}\}) = \frac{\nabla_{\mathbf{r}_i}\Omega^{\rm s}(\{\mathbf{r}\})}{\Omega^{\rm s}(\{\mathbf{r}\})} \qquad \text{(S36)}$$

so that Eq. S32 becomes

$$\begin{aligned}\beta\mathbf{f}_{\mathbf{r}_i} &= \sum_{j=1}^{N_\mathrm{p}}\left[\frac{\nabla_{\mathbf{r}_i}\Omega_j(\{\mathbf{r}\})}{\Omega_j(\{\mathbf{r}\})}\left(\overline{n}_j^{\mathrm{A}}+\overline{n}_j^{\mathrm{B}}+\overline{n}_{jj}^{\mathrm{AB}}\right)+\frac{\nabla_{\mathbf{r}_i}\Omega_j^{\mathrm{s}}(\{\mathbf{r}\})}{\Omega_j^{\mathrm{s}}(\{\mathbf{r}\})}\overline{n}_j^{\mathrm{AC}}\right]\\ &+\sum_{1\le j<q\le N_\mathrm{p}}\frac{\nabla_{\mathbf{r}_i}\Omega_{jq}(\{\mathbf{r}\})}{\Omega_{jq}(\{\mathbf{r}\})}\left(\overline{n}_{jq}^{\mathrm{AB}}+\overline{n}_{jq}^{\mathrm{BA}}\right)+\overline{n}_{\mathrm{s}}^{\mathrm{C}}\frac{\nabla_{\mathbf{r}_i}\Omega^{\mathrm{s}}(\{\mathbf{r}\})}{\Omega^{\mathrm{s}}(\{\mathbf{r}\})}-\beta\nabla_{\mathbf{r}_i}V(\{\mathbf{r}\})\end{aligned}\tag{S37}$$

Using Main Eqs. 11, along with the fact that Ω_j^{s} and Ω_{jq} are only function, respectively, of $\mathbf{r}_j$ and $\{\mathbf{r}_j,\,\mathbf{r}_q\}$ we find

$$\begin{aligned}\beta\mathbf{f}_i &= \sum_{j\in v(i)}\left[\left(\overline{n}_{ij}^{\mathrm{AB}}+\overline{n}_{ij}^{\mathrm{BA}}\right)\frac{\nabla_{\mathbf{r}_i}\Omega_{ij}(r_{ij})}{\Omega_{ij}(r_{ij})}-\left(\overline{n}_j^{\mathrm{A}}+\overline{n}_j^{\mathrm{B}}+\overline{n}_{jj}^{\mathrm{AB}}\right)\frac{\nabla_{\mathbf{r}_i}e_{ji}(r_{ij})}{\Omega_j(\{\mathbf{r}\})}\right.\\ &\left.-\left(\overline{n}_i^{\mathrm{A}}+\overline{n}_i^{\mathrm{B}}+\overline{n}_{ii}^{\mathrm{AB}}\right)\frac{\nabla_{\mathbf{r}_i}e_{ij}(r_{ij})}{\Omega_i(\{\mathbf{r}\})}\right]+\overline{n}_i^{\mathrm{AC}}\frac{\nabla_{\mathbf{r}_i}\Omega_i^{\mathrm{s}}(r_{i,z})}{w_i^{\mathrm{s}}(r_{i,z})}\\ &-\overline{n}_{\mathrm{s}}^{\mathrm{C}}\frac{\nabla_{\mathbf{r}_i}k_i^{\mathrm{s}}(r_{i,z})}{\Omega^{\mathrm{s}}(\{\mathbf{r}\})}-\left(\overline{n}_i^{\mathrm{A}}+\overline{n}_i^{\mathrm{B}}+\overline{n}_{ii}^{\mathrm{AB}}\right)\frac{\nabla_{\mathbf{r}_i}e_i^{\mathrm{s}}(r_{i,z})}{\Omega_i(\{\mathbf{r}\})}-\beta\nabla_i V(\{\mathbf{r}\})\end{aligned}\tag{S38}$$

When using the EXP algorithm (see Main Sec. 2.2), the linkages are evolved using the Gillespie algorithm (see next section). In this case, $\mathbf{f}_i$ is calculated as in the last equality of Eq. S31 in view of the fact that $\{n\}$ are not a function of $\{\mathbf{r}\}$. Therefore, the expression of $\mathbf{f}_i$ used in the EXP case is identical to Eq. S38 when replacing $\{\overline{n}\}$ with the actual values of the linkages $\{n\}$.

Concerning the grand-canonical Monte Carlo algorithm, insertion/removal acceptances are given by

$$\begin{aligned}\mathrm{acc}_{\mathrm{ins}} &= \min\left[1,\frac{V}{(N_\mathrm{p}+1)}\cdot\rho_{\mathrm{id}}\cdot\exp\left[-\beta\Delta\mathcal{F}_{\mathrm{ins}}\right]\right]\\ \mathrm{acc}_{\mathrm{rem}} &= \min\left[1,\frac{N_\mathrm{p}}{V}\cdot\frac{1}{\rho_{\mathrm{id}}}\cdot\exp\left[-\beta\Delta\mathcal{F}_{\mathrm{rem}}\right]\right]\end{aligned}\tag{S39}$$

where $\Delta\mathcal{F}_{\mathrm{rem}}/\Delta\mathcal{F}_{\mathrm{ins}}$ is the change of the system free energy after removing/inserting a colloid in the simulation box. To optimize the acceptances, we set $\Delta\mathcal{F}_{\mathrm{rem}}$ and $\Delta\mathcal{F}_{\mathrm{ins}}$ to zero when the inserted/removed particle does not interact with any other particle or the surface.

S3.1 Gillespie algorithm

In this section, we detail the implementation of the Gillespie algorithm that has been used to simulate sticky-ends reactions in the EXP method.

At a given colloid configuration, $\{\mathbf{r}\}$, we start calculating all on/off rates of making/breaking linkages

$$k_{\rm on}^{ij}, \qquad k_{\rm on}^{ii}, \qquad k_{\rm on}^{i,{\rm s}}, \qquad k_{\rm off}^{ij}, \qquad k_{\rm off}^{ii}, \qquad k_{\rm off}^{i,{\rm s}},$$

as derived in Main Eq. 8. Notice that, for instance, $k_{\rm on}^{ij}$ is the on rate of making a bridge between i and j either using an A sticky end tethered to particle i or j. While off rates are only function of ΔG_0 or $\Delta G_0^{\rm s}$, on rates also include configurational terms that are function of $\{\mathbf{r}\}$ (see Main Eq. 8). Accordingly, the list of all possible reactions is specified by the following affinities

$$\begin{aligned} a_{\rm on,AB}^{ij} = n_i^{\rm A} n_j^{\rm B} k_{\rm on}^{ij}, \quad a_{\rm on,BA}^{ij} = n_i^{\rm B} n_j^{\rm A} k_{\rm on}^{ij}, \quad a_{\rm on}^{ii} = n_i^{\rm A} n_i^{\rm B} k_{\rm on}^{ii}, \quad a_{\rm on}^{i,{\rm s}} = n_i^{\rm A} n_{\rm s}^{\rm C} k_{\rm on}^{i,{\rm s}}, \\ a_{\rm off,AB}^{ij} = n_{ij}^{\rm AB} k_{\rm off}^{ij}, \quad a_{\rm off,BA}^{ij} = n_{ij}^{\rm BA} k_{\rm off}^{ij}, \quad a_{\rm off}^{ii} = n_{ii} k_{\rm off}^{ii}, \quad a_{\rm off}^{i,{\rm s}} = n_i^{\rm AC} k_{\rm off}^{i,{\rm s}}, \end{aligned} \tag{S40}$$

where, for instance, $a_{\rm on,AB}^{ij}$ refers to the possibility of forming a linkage between i and j using an A sticky end tethered to particle i. We then fire one within all possible reactions with probability

$$\begin{aligned} p_{\rm on,AB}^{ij} = \frac{n_i^{\rm A} n_j^{\rm B} k_{\rm on}^{ij}}{a_{\rm tot}}, \quad p_{\rm on,BA}^{ij} = \frac{n_i^{\rm B} n_j^{\rm A} k_{\rm on}^{ij}}{a_{\rm tot}}, \quad p_{\rm on}^{ii} = \frac{n_i^{\rm A} n_i^{\rm B} k_{\rm on}^{ii}}{a_{\rm tot}}, \quad p_{\rm on}^{i,{\rm s}} = \frac{n_i^{\rm A} n_{\rm s}^{\rm C} k_{\rm on}^{i,{\rm s}}}{a_{\rm tot}}, \\ p_{\rm off,AB}^{ij} = \frac{n_{ij}^{\rm AB} k_{\rm off}^{ij}}{a_{\rm tot}}, \quad p_{\rm off,BA}^{ij} = \frac{n_{ij}^{\rm BA} k_{\rm off}^{ij}}{a_{\rm tot}}, \quad p_{\rm off}^{ii} = \frac{n_{ii} k_{\rm off}^{ii}}{a_{\rm tot}}, \quad p_{\rm off}^{i,{\rm s}} = \frac{n_i^{\rm AC} k_{\rm off}^{i,{\rm s}}}{a_{\rm tot}}, \end{aligned} \tag{S41}$$

where $a_{\rm tot}$ is the total affinity. Along with the type of reaction we sample the time for it to happen (τ), distributed as $P(\tau) = \exp[-\tau/a_{\rm tot}]/a_{\rm tot}$, and increment a reaction clock $\tau_{\rm reac}$ by τ. If $\tau_{\rm reac} < \Delta t$ (where Δt is the simulation step), we update $\{n\}$, recalculate the affinities

(Eq. S40), and fire a new reaction until reaching Δt.