Density-tunable pathway complexity in a minimalistic self-assembly model

Soft Matter

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Density-tunable pathway complexity in a minimalistic self-assembly model

Matteo Becchi,[a] Riccardo Capelli,[b,c] Claudio Perego,[d] Giovanni M. Pavan,*[b,d] and Cristian Micheletti*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

An open challenge in self-assembly is learning how to design systems that can be conditionally guided towards different target structures depending on externally-controlled conditions. Using a theoretical and numerical approach, here we discuss a minimalistic self-assembly model that can be steered towards different types of ordered constructs at the equilibrium by solely tuning a facile selection parameter, namely the density of building blocks. Metadynamics and Langevin dynamics simulations allow us to explore the behavior of the system in and out of equilibrium conditions. We show that the density-driven tunability is encoded in the pathway complexity of the system, and specifically in the competition between two different main self-assembly routes. A comprehensive set of simulations provides insight into key factors allowing to make one self-assembling pathway prevailing on the other (or *vice versa*), determining the selection of the final self-assembled products. We formulate and validate a practical criterion for checking whether a specific molecular design is predisposed for such density-driven tunability of the products, thus offering a new, broader perspective to realize and harness this facile extrinsic control of conditional self-assembly.

1 Introduction

Self-assembling systems, where fundamental building blocks, or monomers, interact with each other and organize spontaneously into supramolecular structures, offer promising routes to build new types of materials with tunable properties [1–6]. Notable natural examples include protein filaments [7], tubules [8], viral capsids [9,10], and amyloid fibrils [11], only to name a few. In recent years, the fascinating dynamic properties of such natural self-assembled systems motivated considerable efforts in the supramolecular chemistry community, with the aim to synthesize bioinspired artificial systems *via* similar self-assembly principles [4,12,13]. A variety of, *e.g.*, supramolecular polymers [14,15], tubules [16], micelles [17], vescicles [18], knots [19], catenanes [20,21], etc., has been reported, and the factors that control their self-assembly have been investigated in great detail [22–28].

Within this rapidly advancing frontier, an open challenge is learning how to design self-assembling systems so that they can be externally guided towards different target structures [29]. To this end, a recognised crucial factor for controlling assembly products both in and out of equilibrium [27,30,31] is the so-called "pathway complexity", that is the presence of multiple intermediate states and competing self-assembly routes [30,32]. These defining features have been leveraged to alternatively establish different types of assemblies by modulating intrinsic properties, such as the shape of monomers [33,34]and their interactions [35,36], or by mixing different types of monomers [30,37–39], or using fuel-driven/regulated [31,40,41] and other out-of-equilibrium setups [42,43].

Examples of tuning the pathway complexity *via* extrinsic approaches have been also reported. For example, modifying the temperature [29,44,45], the solvent [29,46–48], or the dilution strategy [21,49] were demonstrated to have an impact on the selection of the self-assembly products. While these extrinsic parameters expectedly have direct bearings on the assembly thermodynamics, these approaches provide examples of how one can externally move the process toward alternative target products. The potential generality of selective assembly strategies based on extrinsic, and hence transferable, control parameters, make them particularly appealing and significant.

Indeed, there are several contexts where qualitatively different assemblies can be obtained with same building blocks at different concentrations, such as micellar versus lamellar structures [50], BCC versus FCC lattices [51] or different non-crystalline phases [52,53].

[a] *Scuola Internazionale Superiore di Studi Avanzati - SISSA, via Bonomea 265, 34136 Trieste, Italy. Email: cristian.micheletti@sissa.it*
[b] *Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy. Email: giovanni.pavan@polito.it*
[c] *Department of Biosciences, Università degli Studi di Milano, Via Giovanni Celoria 26, 20133 Milano, Italy*
[d] *Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, Polo Universitario Lugano, Campus Est, Via la Santa 1, 6962 Lugano-Viganello, Switzerland*

† Electronic Supplementary Information (ESI) available: Supplementary Methods, Supplementary Figures. See DOI: 10.1039/cXsm00000x/

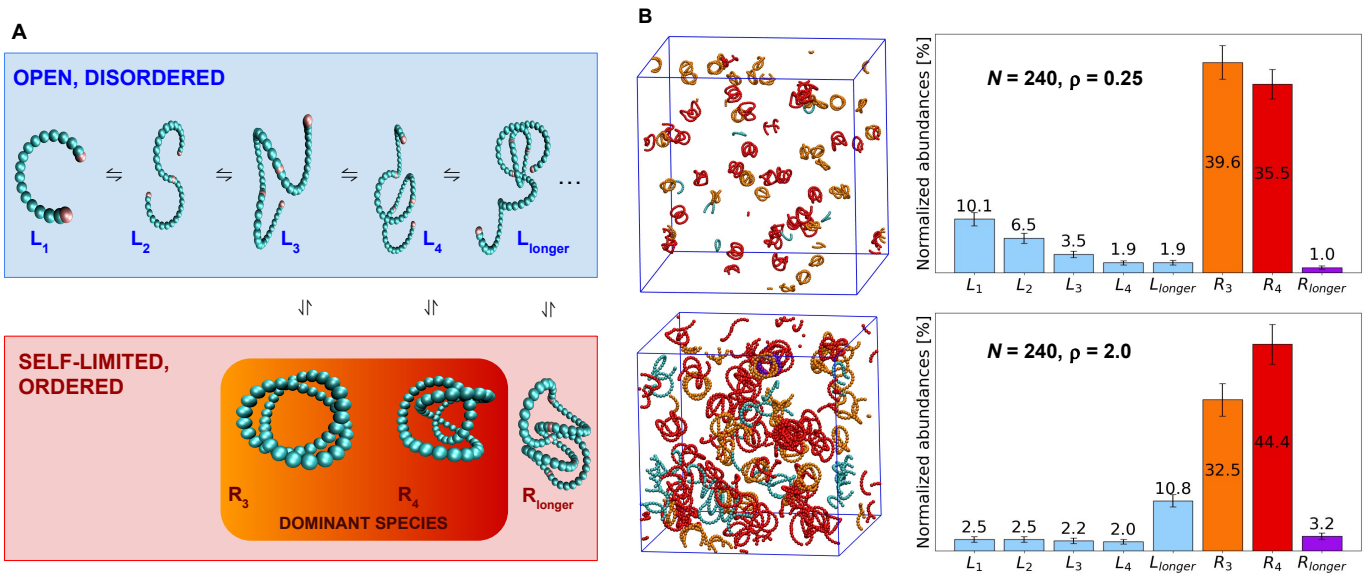

Fig. 1 **Tunability of prevalent assembly products.** (A) The helical building blocks, which are rigid and can bind at the ends *via* attractive patchy interactions (patches are rendered as pink beads; their size is exaggerated for visual clarity), can form linear amorphous constructs (L_n) as well as self-limited ordered rings (R_n), including those highlighted and made of $n = 3$ and $n = 4$ building blocks. For simplicity, populations of linear and circular species with more than 4 building blocks are consolidated in two groups labelled as L_{longer} and R_{longer}, respectively. (B) Equilibrium configuration of systems with $N = 240$ building blocks in periodic cubic simulation boxes with different volumes, V, corresponding to building block densities $\rho = N/V$ equal to 0.25 and 2.0, respectively, expressed in units of a reference concentration $\rho_0 = 7.5 \cdot 10^{-3}\sigma^{-3}$. The histograms show the relative abundances of the different L_n and R_n species at equilibrium and show that varying ρ allows for selecting either R_3 or R_4 as the prevalent assembly type at equilibrium. Histograms with the instantaneous populations of the two snapshots are instead provided in Fig. S1 in the ESI†.

These and analogous instances typically harness the coexistence of distinct bulk phases, involving structures that can extend indefinitely. By contrast, using monomer concentration to select different products among self-limited assemblies has proved much more elusive, and only a few specific examples have been reported, showing the potential of the approach [54].

The inherent challenges for such use of monomer concentration are aptly illustrated by processes resulting in one-dimensional, polymer-like assemblies, such as tubules [55], cylindrical micelles [56] and supramolecular polymers [14,54]. The string-like species formed in equilibrium are poly-dispersed in length, their abundance decaying exponentially with the polymerization degree. Varying monomer density can change the exponential decay length, skewing the population towards shorter or longer species, but no tuning of the density can *e.g.* make a given species more abundant than shorter ones. In addition, varying monomer concentrations can lead to kinetic trapping, disordered aggregation or precipitation. Thus, in most instances it remains prohibitively difficult to obtain clear knowledge on how to design effective one-dimensional systems where selective self-assembly can be achieved by tuning the pathway complexity *via* monomer density.

Here, we introduce and study a minimalistic self-assembly model with an intrinsic pathway complexity, and show unambiguously that it can be selectively guided towards one of two possible target products by solely varying building block concentration. By means of this model, we can thus investigate what features of pathway complexity ultimately underpins the assembly tunability in and out of equilibrium. This also allows us to formulate a practical criterion for checking whether such density-based tunability of the products is intrinsically encoded in a specific molecular design, which ought to be useful to screen candidate systems for conditional self-assembling properties.

2 Methods

2.1 Model

We consider monodispersed mixtures of N patchy helical building blocks, that are treated as rigid bodies and are modelled as in ref. [24,57,58]. The rigid building blocks, that are inspired by synthetic helicates used in supramolecular assembly, consist of $n = 16$ touching spherical beads of diameter σ with two attractive patches at the ends. The centers of the beads lie on a helical curve, parametrised as

$$\vec{r}_i = R(\cos(\alpha i/n), \sin(\alpha i/n), hi/n)$$

where $i \in \{1, \ldots, 16\}$ is the bead index and $R = 3.36\,\sigma$, $\alpha = 1.55\pi$ and $h = 1.008\,R$. The interaction centers of the patches also lie on the helical centerline, close to the outer surface of the terminal beads, see Fig. 1A. The building blocks interact *via* the hard-core repulsion of the beads and the short-range attraction of the patches. The combination of repulsive and short-ranged attractive interactions confers an effective directionality to the latter, which allows one patchy terminal to bind another one, but no more. Self-assembly of branched structures is thus avoided [24,58].

The repulsion is provided by a truncated and shifted (WCA) Lennard-Jones potential,

$$U_{\mathrm{WCA}} = \begin{cases} 4\varepsilon A\left[\left(\frac{\sigma}{d}\right)^{12} - \left(\frac{\sigma}{d}\right)^{6} + \frac{1}{4}\right] & \text{for } d \leq 2^{1/6}\sigma \\ 0 & \text{otherwise} \end{cases}$$

where d is the distance between beads' centers; ε, the characteristic Lennard-Jones energy scale, is customarily taken equal to the system thermal energy $k_{\mathrm{B}}T$.

The attractive, binding potential is instead

$$U_{\mathrm{b}} = \begin{cases} -25\varepsilon B \exp\left[-\frac{d^2}{2\,\tilde{\sigma}^2}\right] & \text{for } d \leq 2^{1/6}\sigma \\ 0 & \text{otherwise} \end{cases}$$

where d is the distance between the centers of the patches and $\tilde{\sigma} = \sigma/10$.

The amplitudes of the potentials are set to their default values, $A = 150$ and $B = 1$ [24,58]. For the out-of-equilibrium setup we used $B = 1.2$, which makes binding practically irreversible. With the default parametrization, bound configurations correspond to distances of about $0.1\,\sigma$ or less. Accordingly, we used a tolerant distance threshold $0.4\,\sigma$ for the algorithmic identification of bound building blocks, see Fig. S2 in the ESI†.

2.2 Langevin Dynamics

Langevin dynamics simulations of the assembly process were carried out for a periodic cubic box of side $L_{\mathrm{b}} = \left(\frac{N}{\rho}\right)^{1/3}$, where ρ is the total building block concentration, expressed in units of a reference concentration $\rho_0 = 7.5 \cdot 10^{-3}\sigma^{-3}$, at which the average nearest neighbor distance is comparable with the gyration diameter of a building block ($\approx 7\,\sigma$). The system initial building block density ρ was varied in the range $0.125 \leq \frac{\rho}{\rho_0} \leq 2.0$.

The molecular dynamics simulations were integrated with the LAMMPS simulation package [59], with unit values, in simulation units, for the friction coefficient, γ, and the beads' mass, m. The integration time step was set equal to $0.012\,\tau_{\mathrm{LJ}}$, where $\tau_{\mathrm{LJ}} = \sigma\sqrt{m/\varepsilon}$ is the characteristic Lennard-Jones time.

We simulated $N = 36$ and $N = 240$ building blocks and, for each combination of ρ and N, we collected from 20 to 40 independent trajectories of duration between $24 \cdot 10^6$ and $120 \cdot 10^6\,\tau_{\mathrm{LJ}}$.

For the default amplitude of the attractive potential, $B = 1$, the lifetime of dimers is about $480 \cdot 10^3\,\tau_{\mathrm{LJ}}$ [58], much larger than the typical self-diffusion time of the building blocks, $\tau_{\mathrm{B}} \approx 20\,\tau_{\mathrm{LJ}}$ [24]. The characteristic binding energy, ε_{b}, was estimated by linearly fitting the average potential energies of different assemblies (dimers, trimers, R_3 and R_4) versus the number of bonds, see Section 1.2 and Fig. S4 in the ESI†.

2.3 Free-energy and kinetics calculations

The free-energy landscapes associated to the $N = 5$ building blocks system at different densities were characterised by means of multiple-walkers (96 replicas) well-tempered metadynamics (WT-MetaD) [60], which enhances the exploration of the conformational space with a history-dependent bias potential along a given set of order parameters, the collective variables (CVs). Two CVs were biased during the WT-MetaD simulations, controlling the overall distance and number of established contacts between the building blocks (see details in the Section 1.1 in the ESI†). The initial height of the Gaussians was $0.1\,k_{\mathrm{B}}T$, with bias factor (the damping parameter) set to 30. The free-energy obtained from the exploration of the CV space was then projected *via* a reweighting procedure [61] onto the different aggregation states indicated in Fig. 2 (see also Section 1.1 in the ESI†).

The kinetics of transition between different states was estimated by means of *infrequent* WT-MetaD simulations [62,63]. Operatively, we performed 100 repetitions of each selected state transition, applying a WT-MetaD bias potential to trigger the change of state. The deposition frequency was set to 10^5 steps^{-1} with an initial height of $0.1\,k_{\mathrm{B}}T$, bias factor of 30. The empirical time distributions obtained following ref. [62] were then fitted to a rare-events distribution [63] to estimate the unbiased rates of transitioning to a target state from a *given* initial one, reported in Fig. 2.

The MetaD simulations were performed by coupling the PLUMED 2.6 library [64,65] with LAMMPS [66]. As for the unbiased Langevin dynamics we set an integration time step of $0.012\,\tau_{\mathrm{LJ}}$. The building block density is modulated by setting a different value of the simulation box side L_{b}.

3 Results and discussion

3.1 Overview of density-driven assembly selection

We considered monodispersed solutions of rigid, curved building blocks [24] that can bind at their ends via attractive patchy interactions (see Methods). The helical shape of the building blocks, which was selected with a general design principle discussed further below, allows for the formation of both linear (string-like) and self-limited circular (knotted) structures. Elements of these two families are indicated as L_n and R_n, respectively, where n is the number of building blocks, see Fig. 1A.

Two remarkable properties of this model readily emerge from assembly simulations at different initial building block density $\rho = N/V$, N being the total number of building blocks in the considered volume V, see Fig. 1B. First, the ordered assemblies R_3 and R_4 are prevalent against the background of disordered, open polymers L_n in the entire range of considered densities ρ. Next, R_3 and R_4 can be alternatively selected as the globally dominant species by solely tuning ρ.

In the following we discuss more in detail the factors underpinning the notable properties of the model, whose minimalistic nature make it ideally suited to extensive systematic characterization of the assembly pathways both in and out of equilibrium.

3.2 Free energy landscape

As a first step, we discuss the case of $N = 5$ building blocks, which is the smallest system where finite size effects do not inevitably introduce enthalpic biases in configurations with either R_3 or R_4 constructs. In fact, the lowest energy microstates featuring R_3 and R_4 are $R_3 + L_2$ and $R_4 + L_1$, which have the same total number of bonds between building blocks (four). The same does not occur in other small-sized systems, such as $N = 4$ or $N = 6$.

The limited size of the systems makes it amenable to use metadynamics protocols [60,62] to establish the free-energies of the various microstates as well as the height of the barriers connecting them.

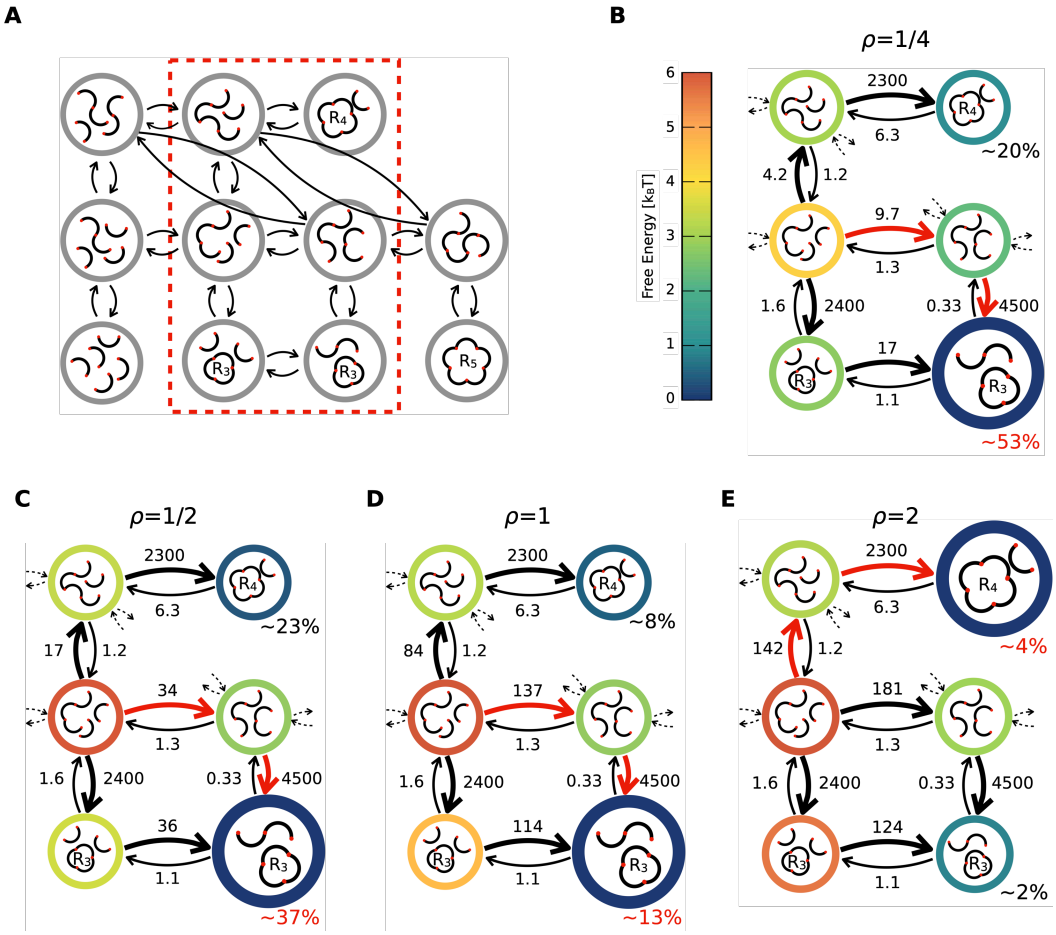

Fig. 2 **Thermodynamics and kinetics of a system composed of 5 building blocks.** (A) Schematic representation of all states (in the circles) and their inter-conversions (arrows) in the system. The red dashed box encloses the key microstates, *i.e.*, those where the prevalent complex assemblies R_3 or R_4 are present, or can be formed *via* a single elementary step. (B-E) Inter-conversion maps between the key states for different building block densities (from $\rho = 0.25$ to $\rho = 2.0$). The colors of the rings enclosing each state represent its free energy in k_BT (color bar). Dashed arrows identify interconnections with the states not shown in the panels. The numbers on the inter-conversion arrows represent the rates of transition from the given initial states to the target ones (in units of 10^9 MD steps^{-1}). For the two dominant ordered states, R_3 and R_4, the equilibrium canonical probabilities at the different concentrations are also reported. Complete results accounting for the entire set of microstates (*i.e.*, also those out of the red dashed box) are provided in the Fig. S3 in the ESI†.

The full set of microstates and possible inter-conversion paths are shown in Fig. 2A. The dashed box encloses the key microstates, *i.e.* those where R_3 or R_4 are present or can be formed in a single elementary step.

Figs. 2B-E show the equilibrium abundances of the key microstates, and the transition rates among them at different building block densities, ρ. Here and in the rest of the text, ρ is expressed in units of a reference concentration $\rho_0 = 7.5 \cdot 10^{-3}\sigma^{-3}$ (see Methods). At all considered ρ's, the fastest processes (thicker arrows) are the growth of linear constructs and linear to circular conversions. Varying the concentration has virtually no impact on dissociation steps or inter-conversions between circular and linear forms, while it strongly modulates the growth rate of linear types. Fig. 2 shows unambiguously that the bias towards growing linear constructs with increasing ρ suffices to change the most probable assembly pathway in the system (marked with red arrows) and to flip the relative balance of R_3 and R_4. This clarifies that the inversion of R_3 and R_4 populations can be rationalised in terms of the system pathway complexity and how it is coupled to the building block concentration, as we discuss below.

3.3 Stochastic dynamics of self-assembly

We now turn to larger systems, from $N = 36$ to $N = 240$ building blocks, for which we study the assembly process with stochastic (Langevin) molecular dynamics simulations. Differently from the $N = 5$ case, where the small system size made it possible to consider microstates involving all the building blocks, the larger systems are more appropriately analysed in terms of the populations or concentrations of the linear and circular species.

The results for a system of $N = 36$ building blocks are presented in Fig. 3, which shows the asymptotic, equilibrium abundance of various species over a 16-fold variation of initial building block density, from $\rho = 0.125$ to $\rho = 2.0$. The data for this larger system confirm the concentration-driven population inversion of the two prevalent species, with the larger R_4 assembly taking over at larger densities. Notice that the curve for R_4 abundance is non-monotonic

and, similarly to the R_3 species, it declines for sufficiently large values of ρ. This is caused by more building blocks being co-opted into longer linear constructs, as in other systems assembled from linear building blocks[67–69]. However, the concentration of each type of linear constructs remains much below those of R_3 and R_4 at all values of ρ. Circular constructs with more than 4 building blocks can be formed too, as illustrated by the R_{longer} assembly in Fig. 1A. In the minimalistic case $N = 5$, these constructs can become prevalent with increasing density, superseding the shorter circular species, R_4 and R_3 see Fig. S3 in the ESI†. However, this does not occur in the considered system, $N = 36$, nor in the larger one of $N = 240$, where R_{longer} constructs remain marginally populated. We note that the species abundances for $N = 36$ and $N = 240$ are all consistent, cf. Figs. 1 and 3, indicating that finite size effects are marginal in these systems.

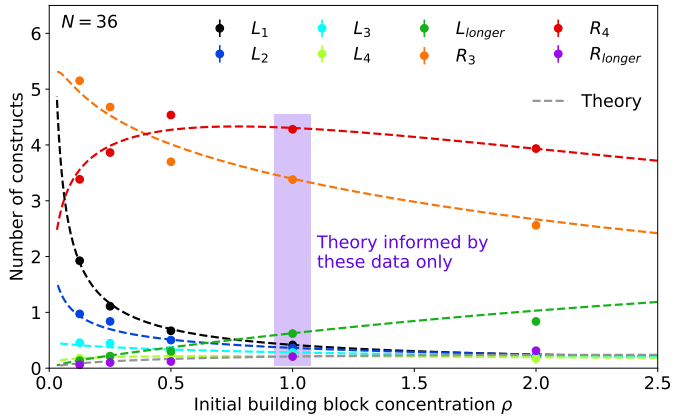

Fig. 3 **Equilibrium populations** Average equilibrium populations of different species, from MD simulations with $N = 36$ building blocks (dots) together with the theoretical prediction (dashed lines), as a function of building block density ρ. The simulation points are averaged over 20 independent simulations starting from different initial configurations; error bars are smaller than the data points. The theoretical predictions are for infinite systems, and are based on parameters fitted from the simulation at $\rho = 1.0$, see Fig. 4. The more abundant species can be modulated varying the building block density above or below a certain crossover concentration located around $\rho_c = 0.397$.

Quantitative insight into the tunable selection of the prevalent assemblies can be gained with the stat-mech framework of cluster formation. In dilute conditions, the equilibrium concentration of species i is

$$\rho_i = \rho_1^n v_i^{n-1} e^{-\beta E_i}, \tag{1}$$

where $\beta = 1/k_B T$, ρ_1 is the (equilibrium) concentration of free building blocks, n is the number of building blocks in the construct, v_i is a (species dependent) characteristic volume and E_i is the binding energy of the species, which is proportional to the number of bonds between building blocks (see Section 1.2 and Fig. S4 in the ESI†).

The above expression holds when the system density is low enough that multi-body interactions, including steric ones, can be neglected[70,71]. We found *a posteriori* that Eq. 1 provides a viable description of the abundance of both self-limited and linear species in our system, even though the building block density ρ is varied significantly.

For each of our observed species we computed the concentration independent term $v_i^{n-1} e^{-\beta E_i}$ by inverting Eq. 1 and using the asymptotic concentrations ρ_i and ρ_1 evaluated at one single building block density, $\rho = 1.0$, see Fig. 4. With this proviso, Eq. 1 can be used predictively to yield the concentrations of all species of infinite systems at arbitrary densities. To this end it suffices to treat the sole free parameter, ρ_1, as quantity to be set self-consistently so to yield the correct system density (see Section 1.3 in the ESI†).

The predicted concentrations are shown in Fig. 3 as dashed curves. Although, we stress, the latter are for infinite systems and were not fitted to the data points, they closely interpolate the points. The good agreement of the actual and predicted populations shows *a posteriori* that the theory of cluster formation remains applicable even at densities where the nominal distance between building blocks, $L_b/N^{1/3}$, becomes comparable to their characteristic size derived from the second virial coefficient, which we computed following ref.[72]. For instance, at $\rho = 2.0$, the nominal distance is $10.2\,\sigma$ and the characteristic building block size is $6.2\,\sigma$, see Section 1.4 in the ESI†.

The robust match of the observed and predicted concentrations has important practical implications. Namely, it allows for using measurements taken at one sole density to test whether the concentration-driven tunability is encoded into the system or not. This was indeed the criterion that we used to single out the shape of the considered building blocks. In fact, the theoretical curves allow to pinpoint the critical density at which population inversion occurs as $\rho_c = 0.397$. The more general chemical-physical implications are that, provided that measurement at one single concentration can be made on a given system, one can directly establish whether the latter affords control over the density-driven tunability of pathway complexity and over the selection of the products. Moreover, this theoretical scheme also allows to determine the temperature range inside which the density-driven product selection is feasible.

Further insight into the tunability of the assembly outcome with building block density is provided by the diagram of Fig. 5. The Cartesian axes report the yield of the two ordered dominant species, measured as the fraction of involved building blocks $(3N_{R_3} + 4N_{R_4})/N$, and the system composition, measured as the N_{R_4}/N_{R_3} ratio. The diagram provides a synoptic representation of the assembly dynamics at different ρ from the initial condition of $N = 36$ dispersed building blocks to the asymptotic, equilibrium state. The trajectories are overlaid to a curvilinear grid that represents the theoretical predictions of the equilibrium yields and composition at different ρ's and different binding energies (in reduced units), $\beta\varepsilon_b$.

The diagram establishes that the evolution towards equilibrium from the initial dispersed state mostly proceeds along constant-ρ grid lines. As time advances, the average dynamical trace for a given ρ visits composition-yield values pertinent to equilibrium states at the same density but with the magnitude of the binding energy (*i.e.* its absolute value) progressively increasing towards the actual value of the system, $\beta\varepsilon_b = -22.69$ (Section 1.2 and Fig. S4 in the ESI†). This remarkable regularity of the trends in

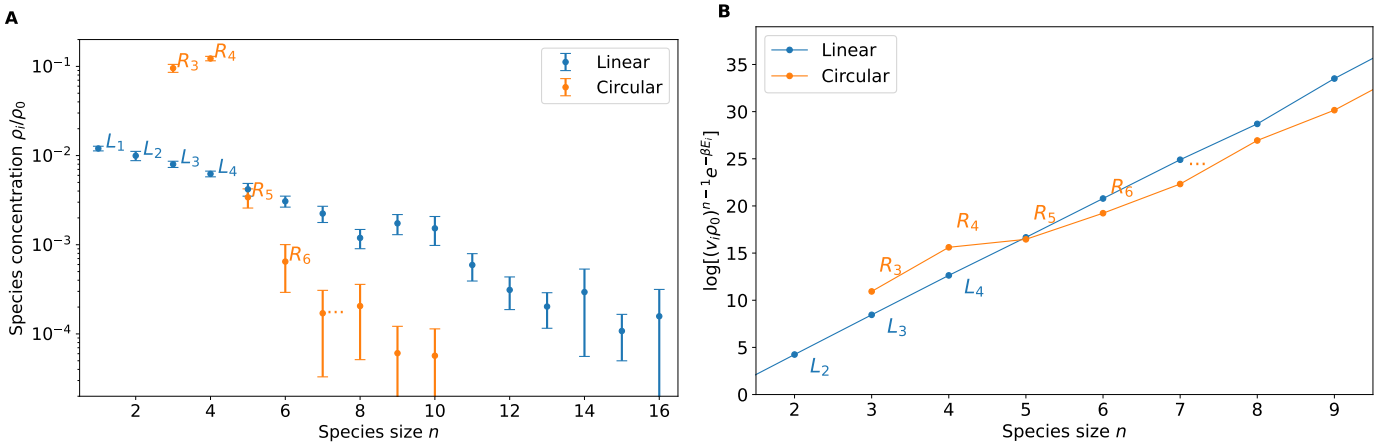

Fig. 4 **Data from the simulation at a single density** $\rho = 1.0$ (A) Concentrations of the different species, linear (blue) and circular (orange), at equilibrium, measured from simulation at $\rho = 1.0$. Data are averages over 20 independent initial configurations. The abundance of open species decays approximately exponentially with the size n. Species R_3 and R_4 dominate over other types of circular and linear assemblies. (B) Logarithm of the weighting factor, $v_i^{n-1}e^{-\beta E_i}$, for a given species i, made of n building blocks. The density-independent weighting factor was calculated from the species' concentrations at $\rho = 1.0$ (panel A) using Eq. 1. The weighting factors were used to obtain the theoretical predictions of the concentrations in Fig. 3 for $\rho \neq 1.0$.

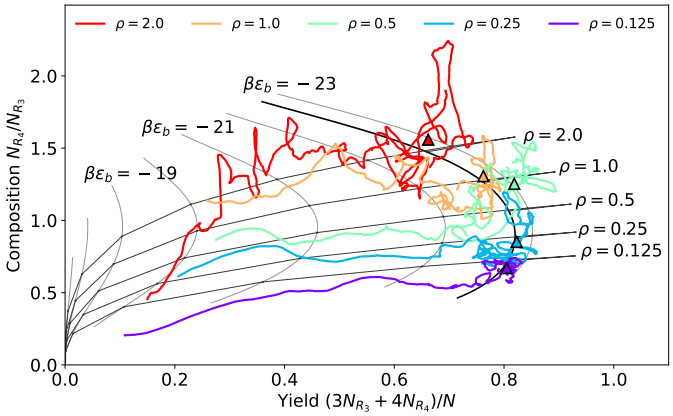

Fig. 5 **Composition-yield diagram.** The plot shows (gray lines) the predicted equilibrium values of the assembly yield $(3N_{R_3} + 4N_{R_4})/N$ and the composition of the assemblies, measured as the ratio between the two main product concentrations N_{R_4}/N_{R_3}, for different values of the building block density ρ (increasing from bottom to top) and the characteristic binding enthalpy $\beta\varepsilon_b$ (increasing from left to right). The black line is the one corresponding to the binding enthalpy $\beta\varepsilon_b = -22.69$ used in the simulations. Colored lines show the path followed on this diagram by the MD simulations, at the different ρ. The starting points of the simulations are on the lower left, and there is a phase of initial relaxation, followed by fluctuations around the equilibrium. The bold triangles show the average values at equilibrium.

Fig. 5 gives quantitative substance to the heuristic conclusion that the system evolution from an initial unbound state is not dissimilar from a quasi-equilibrium assembly process where temperature is gradually lowered from a large value (yielding unbound building blocks) to the actual one.

The observed property is expected to hold more in general[73,74] and, we surmise, might be harnessed to devise informed out-of-equilibrium protocols where the assembly products could be purposely altered or skewed with respect to the equilibrium case[75,76]. One such possible protocol, directly inspired by our setup, could involve the following steps: initial high-temperature equilibration to populate isolated building blocks, temperature quenching to establish the "attractor" value of $\beta\varepsilon_b$, and finally chemical-locking of the assembly products after a specific time delay. Alternatively to the last step, the system could be repeatedly driven through cycles of low and high temperatures/binding energies, a setting that we will directly address further below.

3.4 Density-tunable pathway complexity

We now examine in detail how the conditional selection of assembly products is encoded in the pathway complexity of the system, and specifically in the striking difference of the main assembly pathways above and below the critical density, $\rho_c \approx 0.4$.

Fig. 6 summarises the assembly dynamics for a system of $N = 36$ building blocks at the smallest and largest considered densities, $\rho = 0.125$ and $\rho = 2.0$. For clarity, the populations of all species with up to 4 building blocks are shown individually, while those of larger linear and circular constructs are consolidated in two groups labelled as L_{longer} and R_{longer}, respectively.

The results show that R_3 and R_4 equilibrium populations are reached much more slowly at $\rho = 2.0$ than at $\rho = 0.125$, the relaxation times being approximately $30 \cdot 10^6 \tau_{LJ}$ and $8 \cdot 10^6 \tau_{LJ}$, respectively. From this we conclude that the process is not diffusion-limited, as otherwise it would be faster at higher concentrations[77]. In fact, the characteristic times for building blocks to diffuse over the nominal inter-particle separation is $\approx 300\,\tau_{LJ}$ ($\rho = 0.125$) and $\approx 50\,\tau_{LJ}$ ($\rho = 2.0$), which are orders of magnitude smaller than the relaxation times noted above. The depletion of the population of individual building blocks also occurs over time scales ($\approx 10^4 \tau_{LJ}$) that largely exceed the diffusive ones. Note that the depletion is much faster at $\rho = 2.0$ where, however, the formation of R_3 and R_4 is slower. Furthermore, the balance of short and long species is opposite at the two building block densities: the most abundant

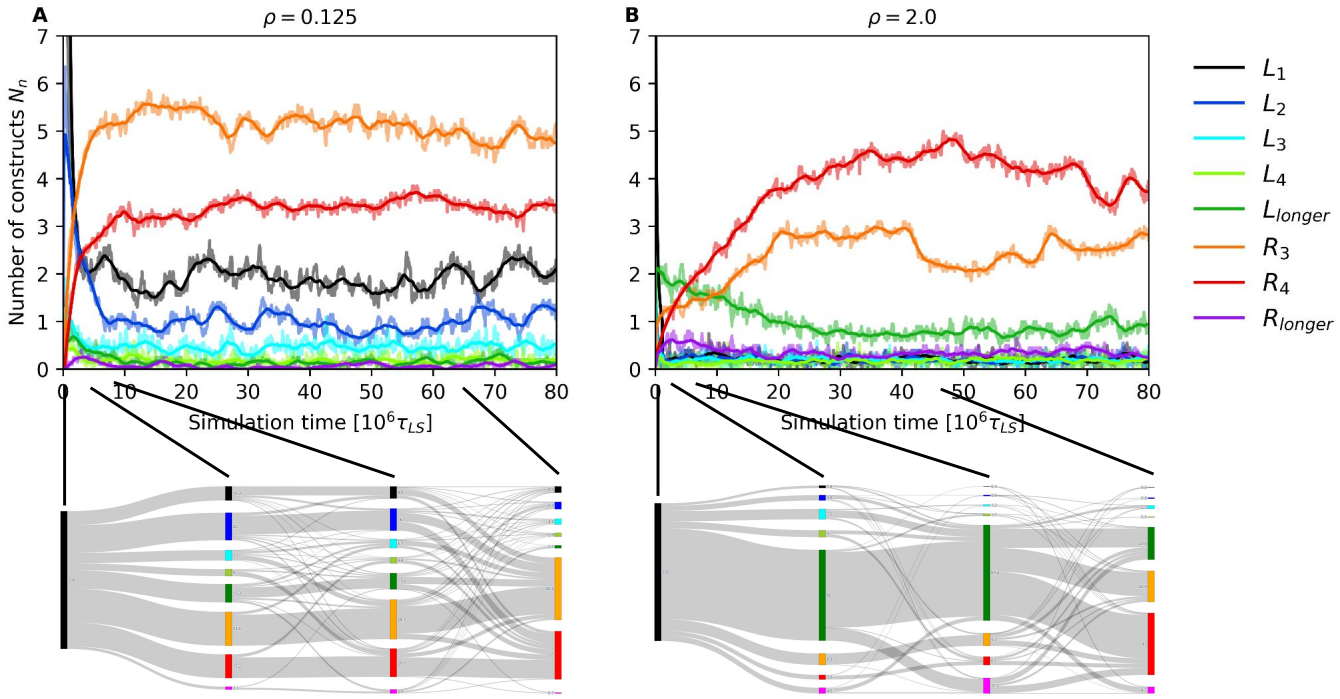

Fig. 6 **Different concentrations select different pathways toward equilibrium.** Pathways leading to the R_3 and R_4 rings (and other species) during the relaxation towards equilibrium, for (A) low ($\rho = 0.125$) and (B) high ($\rho = 2.0$) building block densities (the populations' evolution for intermediate densities is reported in Fig. S5 in the ESI†). In the lower panels, Sankey diagrams show the fluxes among different species for both concentrations, at 4 different time slices: (i) Start of simulation, (ii) peak of linear constructs longer than $n = 4$, (iii) a typical time of bond breaking after the previous slice, and (iv) at equilibrium. The size of each coloured bar is proportional to the total number of building blocks involved in the assemblies of each species. Moreover, the fluxes between different species are shown by the gray bands, having width proportional to the number of building block flowing from one aggregation state to another during each time slice.

types have four or fewer building blocks at $\rho = 0.125$ while they involve three or more building blocks at $\rho = 2.0$.

These diverse aspects are rationalised in the bottom panels of Fig. 6, which show the inter-conversion flow of different species milestoned at four selected times. The milestones correspond to: (i) the initial condition, where only the L_1 unbound building blocks are present, (ii) the maximum population of L_{longer} (iii) a later time delayed by the characteristic lifetime of bonds in linear constructs, $\tau_L \approx 0.5 \cdot 10^6 \tau_{\text{LJ}}$, see Fig. S6 in the ESI†, and finally (iv) the end of the relaxation phase when equilibrium has set in.

Analysis of the flows establishes the following results. At low density, the isolated building blocks bind to form relatively short linear constructs, dimers, trimers and tetramers. Upon closure, the latter two yield the R_3 and R_4 species. Because R_3 and R_4 are formed from the gradual growth and subsequent closure of linear constructs, we dub this pathway as *bottom-up*.

At high density, instead, the isolated building blocks rapidly bind to establish long linear constructs with more than four building blocks. These assemblies later break up into smaller fragments, including L_3 and L_4 species that, again after closure, can accumulate as R_3 and R_4. We dub this pathway as *top-down*.

The emergent behaviour of the system is thus notable in several respects. First, varying the system density introduces a competition between the bottom-up and top-down self-assembly pathways, which are qualitatively different. Next, for increasing density, the top-down pathway antagonizes the bottom-up one, with the result that larger self-limited ordered structures are favoured.

3.5 Harnessing pathway complexity with extreme assembly conditions

The results thus establish building block density as a facile key tuning parameter for selecting the self-assembly outcome by controlling the balance of top-down or bottom-up pathways.

A relevant question is whether there exist other facile assembly conditions or protocols, independent from varying system density, that can be harnessed to further tune the balance between the aforementioned pathways[78]. Identifying such conditions would afford an even higher level of command over the conditional selection of assembly products.

As a specific instance, we considered a general setup where building block binding is periodically switched from reversible to irreversible and *vice versa*. The rationale of the periodic switching is twofold. On the one hand, the irreversible step is introduced to deplete the top-down pathway, which is essentially characterized by dissociation events. On the other hand, the periodic reinstating of binding reversibility is made necessary to avoid runaway linear constructs.

Within our model, this protocol is conveniently realized by varying the building block binding energy at regular intervals, alternating between the default strength and a sufficiently large one (see Methods). In experimental realizations, analogous conditions could be obtained with, *e.g.*, periodic changes of the pH.

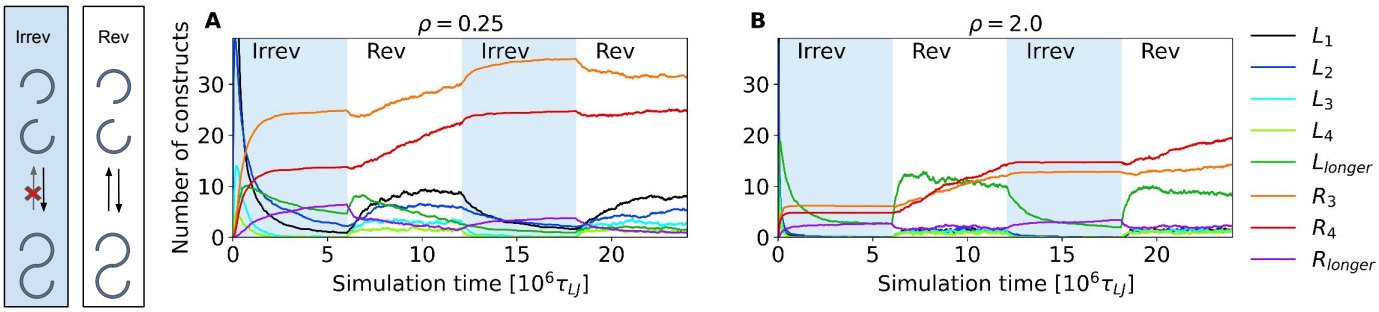

Fig. 7 **Evolution with modulated interaction strength highlights pathway complexity.** Time evolution of the populations of different self-assembled species at (A) low ($\rho = 0.125$) and (B) high ($\rho = 2.0$) building block density ($N = 240$) with the strength of bonding interactions alternatively switched between $B = 1.2$ (irreversible bonding, light blue background) and $B = 1.0$. During the irreversible evolution, one can note a monotonic increase of the rings populations, and a strong interconversion from open (green) to closed (purple) *longer* constructs. Vice versa, during the reversible evolution the *longer* rings break up, increasing the population of *longer* linear constructs, which can subsequently provide building blocks for the R_3 and, mostly, R_4 rings, leading to an increase of the second with respect to the first ones.

Fig. 7 shows the effects of such periodic variations of inter-building block interactions on a system of $N = 240$ building blocks at low and high building block densities ($\rho = 0.25$ and $\rho = 2.0$, respectively). The switching time was set equal to $6 \cdot 10^6 \tau_{\mathrm{LJ}}$, a timescale chosen to be comparable to the characteristic relaxation times of most species populations, see Fig. 6.

At both densities, the irreversible step boosts the formation of long constructs, involving more than 4 building blocks. In fact, the species at the end of the first irreversible phase involve 64 building blocks on average, with some constructs consisting of as many as 213 building blocks. The emergence of such atypically long constructs is a vivid consequence of the suppression of the top-down pathway (where the assemblies can dissociate). In the subsequent reversible phase, the long constructs break-up into shorter fragments, thus favoring the top-down pathway.

The overall effect is that the irreversible step skews the balance in favour of R_3 species, while the reversible steps swings it back towards R_4 assemblies. As the cyclic evolution progresses, the R_3 and R_4 populations are thus differentially ratcheted up towards the steady-state values.

We stress that the choice of timing of the switches and binding energies of our out-of-equilibrium protocols were based on the observed system relaxation kinetics and assembly pathways. Differently from arbitrary choices of the protocol, this informed parametrization allow for tuning the balance of the dominant species beyond what can be achieved at equilibrium by solely intervening on the building block density.

4 Conclusions

In summary, we considered a minimalistic self-assembling model with innate pathway complexity where alternative types of ordered assemblies can be conditionally selected as the prevalent product against a background of disordered constructs by solely tuning the system density. This density-driven tunability of the dominant assembly outcome is noteworthy in several respects. First, it is uncommon in self-assembling systems, like the considered one, where building blocks have two attractive ends, thus yielding polymeric, or string-like constructs. In such systems, varying the density can typically at best skew the heterogeneous population towards shorter or longer constructs, but with no possibility of changing the ranking of the various species. On the other hand, in most practical contexts increasing building block density is often avoided in self-assembling systems as it is generally associated with aspecific aggregation, entrapment in disordered states, and precipitation. Our study shows, instead, that not only such undesired consequences can be escaped, but that different types of ordered self-assembled products can be selected and made prevalent by acting on building block density.

Analysis of the assembly process from an initial state of dispersed building blocks reveals such density-driven tunability is integrated in the system's pathway complexity. The latter involves two main self-assembly routes, which we dub *bottom-up* and *top-down*. In the former, the two competing types of ordered assemblies are formed by accretion, that is, *via* progressive addition of smaller units. In the latter, the same ordered assemblies are preferentially established *via* the breakage and reconfiguration of long linear-like aggregates (fragmentation and recombination). The dominant type of ordered assembled product is controlled by the competition between the two self-assembling routes, which can be tuned by acting solely on building block concentration. This provides a facile way to externally select alternative self-assembly outcomes in equilibrium. As we showed, resorting to out-of-equilibrium assembly protocols can provide additional tunability of the process.

The unusual tunable properties of our model system were achieved thanks to a general design concept based on a stat-mech framework for cluster formation. The principle was used to rationally customize the building blocks' features, in terms of shape, directionality and range of interaction, so to program the conditional selectability of two ordered products. Importantly, we showed that the same design criterion can also be conveniently used to assess, based on population measurements at one density only, whether a self-organizing system is predisposed to such control with varying density.

In consideration of the minimalistic formulation of our model

and the general purpose of our strategy, we expect that a similar approach can be generally applicable, thus providing a powerful search strategy for identifying, in computationally or experimentally efficient manners, candidate systems amenable to conditional selection of the self-assembly outcome. Possible leads include supramolecular structures based on, *e.g.*, DNA origami[79,80], or rigid organic molecules such as helicates[19,81], that self-assemble into closed self-limiting structures, cages, macrocycles, etc.[82–87]. In general, the physical knowledge that can be gleaned from such molecular models in terms of key factors controlling the emergence of complexity and the selection of species during self-assembly holds a considerable potential towards a systematic exploration of naturally-occurring systems with similar properties, as well as for discovering practical manners to control self-assembly in supramolecular chemistry.

Author Contributions

MB performed and analyzed the MD simulations of the self-assembling systems. RC and CP performed and analyzed the metadynamics simulations. GMP and CM designed the research and supervised the work. All authors contributed to the writing and gave their approval to the final version of this paper.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

G.M.P. acknowledges the support received by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement no. 818776 - DYNAPOL) and by the Swiss National Science Foundation (SNSF Grant IZLIZ2_183336). The authors also acknowledge the computational resources provided by SISSA on the Ulysses Supercomputing Cluster, by Swiss National Supercomputing Center (CSCS) and by CINECA.

References

1 D. Philp and J. F. Stoddart, *Angewandte Chemie International Edition in English*, 1996, **35**, 1154–1196.

2 J. H. van Esch and B. L. Feringa, *Angewandte Chemie International Edition*, 2000, **39**, 2263–2266.

3 R. S. Forgan, J.-P. Sauvage and J. F. Stoddart, *Chemical Reviews*, 2011, **111**, 5434–5464.

4 T. Aida, E. Meijer and S. I. Stupp, *Science*, 2012, **335**, 813–817.

5 J. Zhong, L. Zhang, D. P. August, G. F. S. Whitehead and D. A. Leigh, *Journal of the American Chemical Society*, 2019, **141**, 14249–14256.

6 P. Hashim, J. Bergueiro, E. Meijer and T. Aida, *Progress in Polymer Science*, 2020, **105**, 101250.

7 S. E. Ahnert, J. A. Marsh, H. Hernández, C. V. Robinson and S. A. Teichmann, *Science*, 2015, **350**, aaa2245.

8 M. Dogterom and G. H. Koenderink, *Nat Rev Mol Cell Biol*, 2019, **20**, 38–54.

9 J. D. Perlmutter and M. F. Hagan, *Annual Review of Physical Chemistry*, 2015, **66**, 217–239.

10 R. F. Bruinsma, G. J. L. Wuite and W. H. Roos, *Nature Reviews Physics*, 2021, **3**, 76–91.

11 W. Pulawski, U. Ghoshdastider, V. Andrisano and S. Filipek, *Appl Biochem Biotechnol.*, 2012, **166**, 1626–1643.

12 B. A. Grzybowski and W. T. S. Huck, *Nature Nanotechnology*, 2016, **11**, 585–592.

13 A. Levin, T. A. Hakala, L. Schnaider, G. J. L. Bernardes, E. Gazit and T. P. J. Knowles, *Nat Rev Chem*, 2020, **4**, 615–634.

14 L. Brunsveld, B. J. B. Folmer, E. W. Meijer and R. P. Sijbesma, *Chem. Rev.*, 2001, **101**, 4071–4098.

15 D. van der Zwaag, T. F. de Greef and E. Meijer, *Angewandte Chemie International Edition*, 2015, **54**, 8334–8336.

16 W. A. Hansen and S. D. Khare, *Current Opinion in Structural Biology*, 2020, **63**, 106–114.

17 Y. Lu, J. Lin, L. Wang, L. Zhang and C. Cai, *Chem. Rev.*, 2020, **120**, 4111–4140.

18 A. Joseph, C. Contini, D. Cecchin, S. Nyberg, L. Ruiz-Perez, J. Gaitzsch, G. Fullstone, X. Tian, J. Azizi, J. Preston, G. Volpe and G. Battaglia, *Science Advances*, 2017, **3**, e1700362.

19 S. D. P. Fielden, D. A. Leigh and S. L. Woltering, *Angewandte Chemie International Edition*, 2017, **56**, 11166–11194.

20 T. Yamamoto and Y. Tezuka, *Soft Matter*, 2015, **11**, 7458–7468.

21 S. Datta, Y. Kato, S. Higashiharaguchi, K. Aratsu, A. Isobe, T. Saito, D. D. Prabhu, Y. Kitamoto, M. J. Hollamby, A. J. Smith, R. Dalgliesh, N. Mahmoudi, L. Pesce, C. Perego, G. M. Pavan and S. Yagai, *Nature*, 2020, **583**, 400–405.

22 F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi and C. M. Dobson, *Proceedings of the National Academy of Sciences*, 1999, **96**, 3590–3594.

23 A. J. Olson, Y. H. E. Hu and E. Keinan, *Proceedings of the National Academy of Sciences*, 2007, **104**, 20731–20736.

24 G. Polles, D. Marenduzzo, E. Orlandini and C. Micheletti, *Nature Communications*, 2015, **6**, 6423.

25 D. Bochicchio, M. Salvalaglio and G. M. Pavan, *Nature Communications*, 2017, **8**, 147.

26 L. Zhang, J.-F. Lemonnier, A. Acocella, M. Calvaresi, F. Zerbetto and D. A. Leigh, *Proceedings of the National Academy of Sciences*, 2019, **116**, 2452–2457.

27 M. Wehner and F. Würthner, *Nature Reviews Chemistry*, 2020, **4**, 38–53.

28 A. L. de Marco, D. Bochicchio, A. Gardin, G. Doni and G. M. Pavan, *ACS Nano*, 2021, **15**, 14229–14241.

29 M. F. J. Mabesoone, A. J. Markvoort, M. Banno, T. Yamaguchi, F. Helmich, Y. Naito, E. Yashima, A. R. A. Palmans and E. W. Meijer, *Journal of the American Chemical Society*, 2018, **140**, 7810–7819.

30 P. A. Korevaar, S. J. George, A. J. Markvoort, M. M. J. Smulders, P. A. J. Hilbers, A. P. H. J. Schenning, T. F. A. De Greef and E. W. Meijer, *Nature*, 2012, **481**, 492–496.

31 J. Deng and A. Walther, *Journal of the American Chemical Society*, 2020, **142**, 685–689.

32 P. A. Korevaar, T. F. A. de Greef and E. W. Meijer, *Chemistry of Materials*, 2014, **26**, 576–586.

33 S. Ogi, T. Fukui, M. L. Jue, M. Takeuchi and K. Sugiyasu,

Angewandte Chemie International Edition, 2014, **53**, 14363–14367.

34 N. M. Casellas, S. Pujals, D. Bochicchio, G. M. Pavan, T. Torres, L. Albertazzi and M. García-Iglesias, *Chemical Communications*, 2018, **54**, 4112–4115.

35 J. Madge and M. A. Miller, *The Journal of Chemical Physics*, 2015, **143**, 044905.

36 J. D. Halverson and A. V. Tkachenko, *The Journal of Chemical Physics*, 2016, **144**, 094903.

37 B. Adelizzi, A. Aloi, A. J. Markvoort, H. M. M. Ten Eikelder, I. K. Voets, A. R. A. Palmans and E. W. Meijer, *Journal of the American Chemical Society*, 2018, **140**, 7168–7175.

38 A. Sarkar, R. Sasmal, C. Empereur-mot, D. Bochicchio, S. V. K. Kompella, K. Sharma, S. Dhiman, B. Sundaram, S. S. Agasti, G. M. Pavan and S. J. George, *Journal of the American Chemical Society*, 2020, **142**, 7606–7617.

39 A. Murugan, Z. Zeravcic, M. P. Brenner and S. Leibler, *Proceedings of the National Academy of Sciences*, 2015, **112**, 54–59.

40 A. Mishra, D. B. Korlepara, M. Kumar, A. Jain, N. Jonnalagadda, K. K. Bejagam, S. Balasubramanian and S. J. George, *Nature communications*, 2018, **9**, 1295.

41 C. Perego, L. Pesce, R. Capelli, S. J. George and G. M. Pavan, *ChemSystemsChem*, 2021, **3**, e2000038.

42 G. Bisker and J. L. England, *Proceedings of the National Academy of Sciences*, 2018, **115**, E10531–E10538.

43 A. Ben-Ari, L. Ben-Ari and G. Bisker, *The Journal of Chemical Physics*, 2021, **155**, 234113.

44 A. T. Haedler, S. C. J. Meskers, R. H. Zha, M. Kivala, H.-W. Schmidt and E. W. Meijer, *Journal of the American Chemical Society*, 2016, **138**, 10539–10545.

45 K. Venkata Rao, D. Miyajima, A. Nihonyanagi and T. Aida, *Nature Chemistry*, 2017, **9**, 1133–1139.

46 P. A. Korevaar, C. Schaefer, T. F. A. de Greef and E. W. Meijer, *Journal of the American Chemical Society*, 2012, **134**, 13482–13491.

47 J. S. Valera, R. Gómez and L. Sánchez, *Small*, 2018, **14**, 1702437.

48 G. Ghosh and S. Ghosh, *Chemical Communications*, 2018, **54**, 5720–5723.

49 E. Weyandt, L. Leanza, R. Capelli, G. M. Pavan, G. Vantomme and E. W. Meijer, *Nature Communications*, 2022, **13**, 248.

50 F. Sciortino, A. Giacometti and G. Pastore, *Physical review letters*, 2009, **103**, 237801.

51 E. Noya, C. Vega, J. Doye and A. Louis, *The Journal of chemical physics*, 2007, **127**, 054501.

52 H. Pattabhiraman, A. P. Gantapara and M. Dijkstra, *The Journal of chemical physics*, 2015, **143**, 164905.

53 A. Reinhardt, A. J. Williamson, J. P. Doye, J. Carrete, L. M. Varela and A. A. Louis, *The Journal of chemical physics*, 2011, **134**, 104905.

54 L. Su, J. Mosquera, M. F. J. Mabesoone, S. M. C. Schoenmakers, C. Muller, M. E. J. Vleugels, S. Dhiman, S. Wijker, A. R. A. Palmans and E. W. Meijer, *Science*, 2022, **377**, 213–218.

55 M. Kirschner and T. Mitchison, *Cell*, 1986, **45**, 329–342.

56 W. M. Gelbart, A. Ben-Shaul and D. Roux, *Micelles, Membranes, Microemulsions and Monolayers*, Springer, New York, 1994.

57 G. Polles, E. Orlandini and C. Micheletti, *ACS Macro Letters*, 2016, **5**, 931–935.

58 M. Marenda, E. Orlandini and C. Micheletti, *Nature Communications*, 2018, **9**, 3051.

59 A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. Michael Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, *Computer Physics Communications*, 2021, **271**, 108171.

60 A. Barducci, G. Bussi and M. Parrinello, *Physical Review Letters*, 2008, **100**, 020603.

61 P. Tiwary and M. Parrinello, *The Journal of Physical Chemistry B*, 2015, **119**, 736–742.

62 P. Tiwary and M. Parrinello, *Physical Review Letters*, 2013, **111**, 230602.

63 M. Salvalaglio, P. Tiwary and M. Parrinello, *Journal of Chemical Theory and Computation*, 2014, **10**, 1420–1425.

64 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Computer Physics Communications*, 2014, **185**, 604–613.

65 M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi *et al.*, *Nature Methods*, 2019, **16**, 670–673.

66 S. Plimpton, *Journal of Computational Physics*, 1995, **117**, 1–19.

67 H. Jacobson and W. H. Stockmayer, *The Journal of chemical physics*, 1950, **18**, 1600–1606.

68 M. Gordon and W. Temple, *Die Makromolekulare Chemie: Macromolecular Chemistry and Physics*, 1972, **152**, 277–289.

69 G. Ercolani, L. Mandolini, P. Mencarelli and S. Roelens, *Journal of the American Chemical Society*, 1993, **115**, 3901–3908.

70 F. Sciortino, *Soft Matter Self Assembly*, 2016, **193**, 1–17.

71 E. Lattuada, D. Caprara, V. Lamberti and F. Sciortino, *Nanoscale*, 2020, **12**, 23003–23012.

72 D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Elsevier, 2001, vol. 1.

73 F. Sciortino and E. Zaccarelli, *Current Opinion in Solid State and Materials Science*, 2011, **15**, 246–253.

74 F. Sciortino and E. Zaccarelli, *Current Opinion in Colloid & Interface Science*, 2017, **30**, 90–96.

75 W. M. Jacobs, A. Reinhardt and D. Frenkel, *Proceedings of the National Academy of Sciences*, 2015, **112**, 6313–6318.

76 J. Lee Tin Wah, C. David, S. Rudiuk, D. Baigl and A. Estevez-Torres, *ACS nano*, 2016, **10**, 1978–1987.

77 W. Poon and M. Haw, *Advances in Colloid and Interface Science*, 1997, **73**, 71–126.

78 F. M. Gartner, I. R. Graf and E. Frey, *Proceedings of the National Academy of Sciences*, 2022, **119**, e2116373119.

79 H. Dietz, S. M. Douglas and W. M. Shih, *Science*, 2009, **325**, 725–730.

80 C. E. Castro, F. Kilchherr, D.-N. Kim, E. L. Shiao, T. Wauer, P. Wortmann, M. Bathe and H. Dietz, *Nature Methods*, 2011, **8**,

221–229.

81 J. J. Danon, A. Krüger, D. A. Leigh, J.-F. Lemonnier, A. J. Stephens, I. J. Vitorica-Yrezabal and S. L. Woltering, *Science*, 2017, **355**, 159–162.

82 M. Yoshizawa, J. K. Klosterman and M. Fujita, *Angewandte Chemie International Edition*, 2009, **48**, 3418–3438.

83 T. R. Cook and P. J. Stang, *Chemical Reviews*, 2015, **115**, 7001–7045.

84 D. Zhang, T. K. Ronson and J. R. Nitschke, *Accounts of Chemical Research*, 2018, **51**, 2423–2436.

85 L. Pesce, C. Perego, A. B. Grommet, R. Klajn and G. M. Pavan, *Journal of the American Chemical Society*, 2020, **142**, 9792–9802.

86 S. Pullen, J. Tessarolo and G. H. Clever, *Chemical Science*, 2021, **12**, 7269–7293.

87 L. Turcani, A. Tarzia, F. T. Szczypiński and K. E. Jelfs, *The Journal of Chemical Physics*, 2021, **154**, 214102.