Electronic Supplementary Material (ESI) for Soft Matter.
This journal is © The Royal Society of Chemistry 2022

Soft Matter

ARTICLE TYPE

Electronic Supplementary Information for: Density-tunable pathway complexity in a minimalistic self-assembly model

Matteo Becchi, Riccardo Capelli, Claudio Perego, Giovanni M. Pavan, and Cristian Micheletti

1 Supplementary Methods

1.1 Free Energy Calculations

The free-energy landscapes associated with the $N = 5$ building blocks system at different concentrations were characterized using well-tempered metadynamics (WT-MetaD) [1]. This enhanced sampling technique applies a history-dependent bias potential to a set of apt reaction coordinates or collective variables, CVs, to enhance the exploration of the conformational space.

Two CVs were biased during the MetaD simulations: the first CV is the radius of gyration of the system, while the second CV is defined as:

$$\xi = -\gamma^{-1} \log \sum_{i<j} \exp(-\gamma d_{ij}),$$

where the sum runs over all pairs of patches in different building blocks, d_{ij} is the distance of patches i and j, and γ is a parameter that we set to 1. With this choice of γ, the sum in the definition of ξ is dominated by the minimum value of d_{ij}. Moreover, these interactions contribute equally when more than one patchy contact is established. Thus, ξ acts both as a smooth counter of the contacts between patchy particles and, when no contact is formed, as a proxy for the minimum distance between patchy particles.

The history-dependent bias was accumulated in the space of these two CVs in the form of Gaussian-shaped kernels of initial height $0.1\,k_BT$ and width equal to 0.1 for the radius of gyration and 0.05 for ξ, deposited every 1000 simulation steps. In WT-MetaD, the height of the Gaussians, which decreases with the sampling, is controlled by the bias factor, which was set to 30. To accelerate convergence and guarantee an adequate exploration of phase space, we employed 96 parallel replicas in each simulation. The free energy obtained from exploring the CV space at different densities, see Fig. S3, was then projected via a suitable reweighting procedure [2] onto the various microstates indicated in Fig. 2 of the main text. The possible states are identified by the combination of (i) the total number of (pairwise) contacts of the patches, n^c_{tot} and (ii) by the number of building blocks with a single contact, n^c_1. This allows for distinguishing all states, except for the case ($n^c_{\text{tot}} = 4$, $n^c_1 = 2$), which maps onto two states, L_5 and $R_3 + L_2$. These states were distinguished with the determinant of the contact matrix.

The transition kinetics was studied using infrequent WT-MetaD simulations [3,4], an approach that reconstructs the unbiased rates of rare transitions connecting two given states, triggered by a MetaD bias potential. The method consists in performing multiple WT-MetaD simulations of the system starting from the initial state, activating the transition to the final state by a slow build-up of the bias potential in such a way that the transition state is not perturbed by the biasing procedure. The CV considered in our estimation is the distance d_{ij} between the patchy particles involved in the state change (as all transitions in Fig. 2 of the main text involve the formation or the rupture of a bond between 2 patchy particles). In the infrequent MetaD simulations, we performed 100 repetitions of each selected transition, biasing the relevant d_{ij} to trigger the transitions. The bias on the distance was increased every 10^5 steps depositing Gaussians of initial height of 0.1 k_BT, bias factor of 30, and width equal to 0.1. The transition times collected for each event were then converted to unbiased times following Ref. [3]. By fitting the resulting distribution with a Poissonian distribution (the expected time distribution for such rare events) [4] we could estimate the unbiased transition rates reported in Fig. 2 of the main text.

We note that to characterize the kinetics of the system, it is not necessary to simulate both the forward (*on*) and backward (*off*) pathways connecting two states, as we can use the free-energy difference ΔG extracted from WT-MetaD calculations Fig. S3) to estimate the dissociation constant $K_d = \exp(-\Delta G/k_BT)$. Once we have the ΔG and one between k_{on} and k_{off} the kinetics is determined via the rela-

tionship $K_d = k_{\rm off}/k_{\rm on}$.

The MetaD simulations were performed by coupling the PLUMED 2.6 library[5,6] with LAMMPS[7]. As for the unbiased Langevin dynamics we set an integration time step of $0.012\,\tau_{\rm LJ}$. The building block density is modulated by setting a different value of the simulation box side $L_{\rm b}$.

1.2 Binding energy

To clarify how the interplay between enthalpy and entropy drives the assembly process, we first measured the enthalpic gain of building block binding, which is also a key quantity for predicting or extrapolating the species equilibrium populations at various conditions.

We considered a system with $N = 4$ building blocks at $\rho = 1.0$ and used the Langevin dynamics setup to measure the total average potential energy. In this small system, the total number of bonds between building blocks that can be established ranges from zero (only free building blocks) to four (a single R_4 ring). The measured values of the potential energy are shown in the box and whisker plot of Fig. S4, as a function of the number of established bonds between building blocks $n_{\rm b}$. Notice that different states could be associated with the same value of $n_{\rm b}$. For instance, there are two states corresponding to $n_{\rm b} = 3$: the open linear L_4, and the closed ring R_3 plus one free building block.

The data points in Fig. S4 are well interpolated by a linear fit, with slope $\varepsilon_{\rm b} = -22.69\,k_{\rm B}T$ yielding the effective binding energy between building blocks that, to a first approximation, we assumed to be identical across all constructs.

1.3 Prediction of equilibrium populations.

In the high dilution limit, the concentration ρ_i of species i in canonical equilibrium at temperature T is given by[8]:

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

where ρ_1 is the equilibrium concentration of free building blocks, n is the number of involved building blocks, $\beta = 1/k_BT$, v_i is the characteristic volume of the species, and E_i is its binding energy. The latter is proportional to the number of bonds in the species, and is thus equal to $E_i = (n-1)\varepsilon_{\rm b}$ for the open species, and to $E_i = n\varepsilon_{\rm b}$ for the closed ones.

This expression can be inverted to obtain the values of v_i from the equilibrium concentrations ρ_i measured from simulations at a single system density ρ. When substituted in Eq. 1, the inferred values of v_i can then be used to obtain theoretical predictions for the equilibrium abundance of species in systems with N building blocks at arbitrary density, ρ. To this end, ρ_1 is best treated as a free parameter to be set self-consistently from the conservation law:

$$\rho = \sum_{n=1} n\rho_{L_n} + \sum_{n=3} n\rho_{R_n}\ . \tag{2}$$

When using such a predictive scheme we capped the summations in Eq. 2 at $N = 16$ for the linear species and $N = 10$ for the closed ones. This is justified by the exponential decay of the concentration shown in Fig. 4A, and *a posteriori* by the convergence of the predictions.

The expression 1 allows predicting the equilibrium populations not only at different densities ρ but also at different interaction strengths, by simply varying $\varepsilon_{\rm b}$.

1.4 Second virial coefficient

To measure the building blocks' second virial coefficient, B_2, we used Widom insertions[9] as implemented in the LAMMPS simulation package[10]. Starting from a simulation box containing one single building block, we performed $M = 2\cdot 10^9$ insertions of a second building block, recording exclusively the inter-building block interaction energy due to the WCA potential, i.e. disregarding the attractive potential of the patches. The second virial coefficient was computed as

$$B_2 = \frac{1}{2}\left(1 - \frac{1}{M}\sum_{i=1}^{M} e^{-\beta\Delta U_i}\right) V_{\rm box}\ , \tag{3}$$

where M is the number of insertions, $\beta = 1/k_{\rm B}T$, U_i is the aforementioned interaction energy of the $i-$th insertion and $V_{\rm box} = 4096\,\sigma^3$ is the volume of the simulation box.

The calculation returned the value $B_2 = (236.5 \pm 0.1)\ \sigma^3$. For comparison, the gyration radius of the building block is $R_{\rm g} = 3.376\ \sigma$, which gives a gyration volume $V_{\rm g} = 161.19\ \sigma^3$.

References

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

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

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

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

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

6 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.

7 S. Plimpton, *Journal of Computational Physics*, 1995, **117**, 1–19.
8 F. Sciortino, *Soft Matter Self Assembly*, 2016, **193**, 1–17.
9 D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Elsevier, 2001, vol. 1.
10 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.
11 M. Marenda, E. Orlandini and C. Micheletti, *Nature Communications*, 2018, **9**, 3051.

2 Supplementary Figures

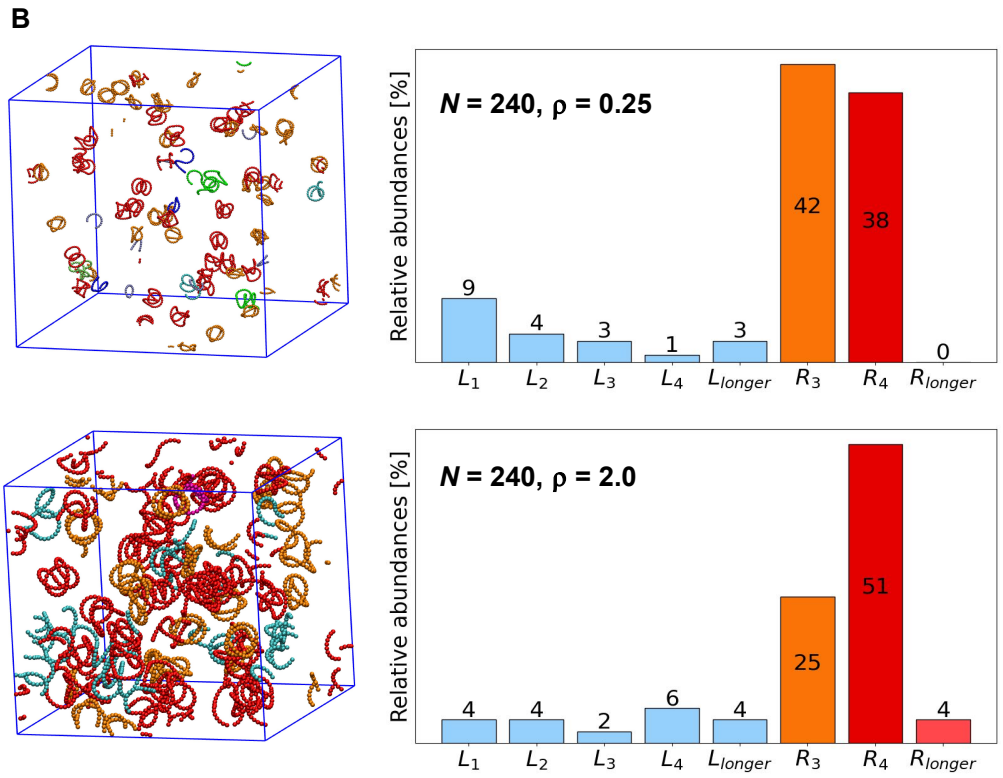

Fig. S 1 **Equilibrium configuration of systems with $N = 240$ building blocks** in periodic cubic simulation boxes with different volumes, V, corresponding to building block densities $\rho = N/V$ equal to 0.25 and 2.0, respectively. The histograms show the instantaneous relative abundances of the different L_n and R_n species and show that varying ρ allows for selecting either R_3 or R_4 as the dominant assembly type at equilibrium.

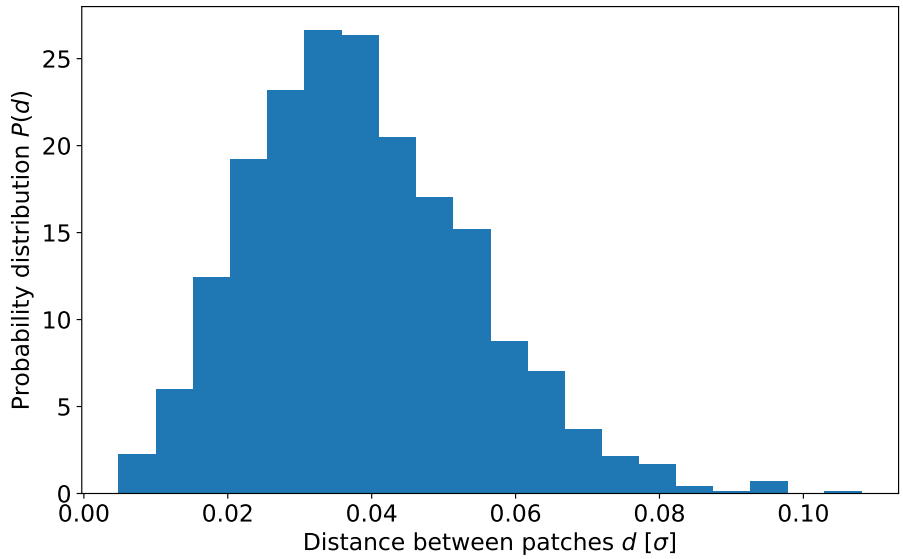

Fig. S 2 **Probability distribution for the distance between bonded patches in a R_4 construct.** The distribution has been obtained from the MD simulation of an assembled R_4 constructs. From this distribution, and also from the SI of ref.[11], it can be seen that the choice of the threshold 0.4 σ is adequate to identify all the bonded building blocks, even in the case that other constructs have slightly broader distance distributions.

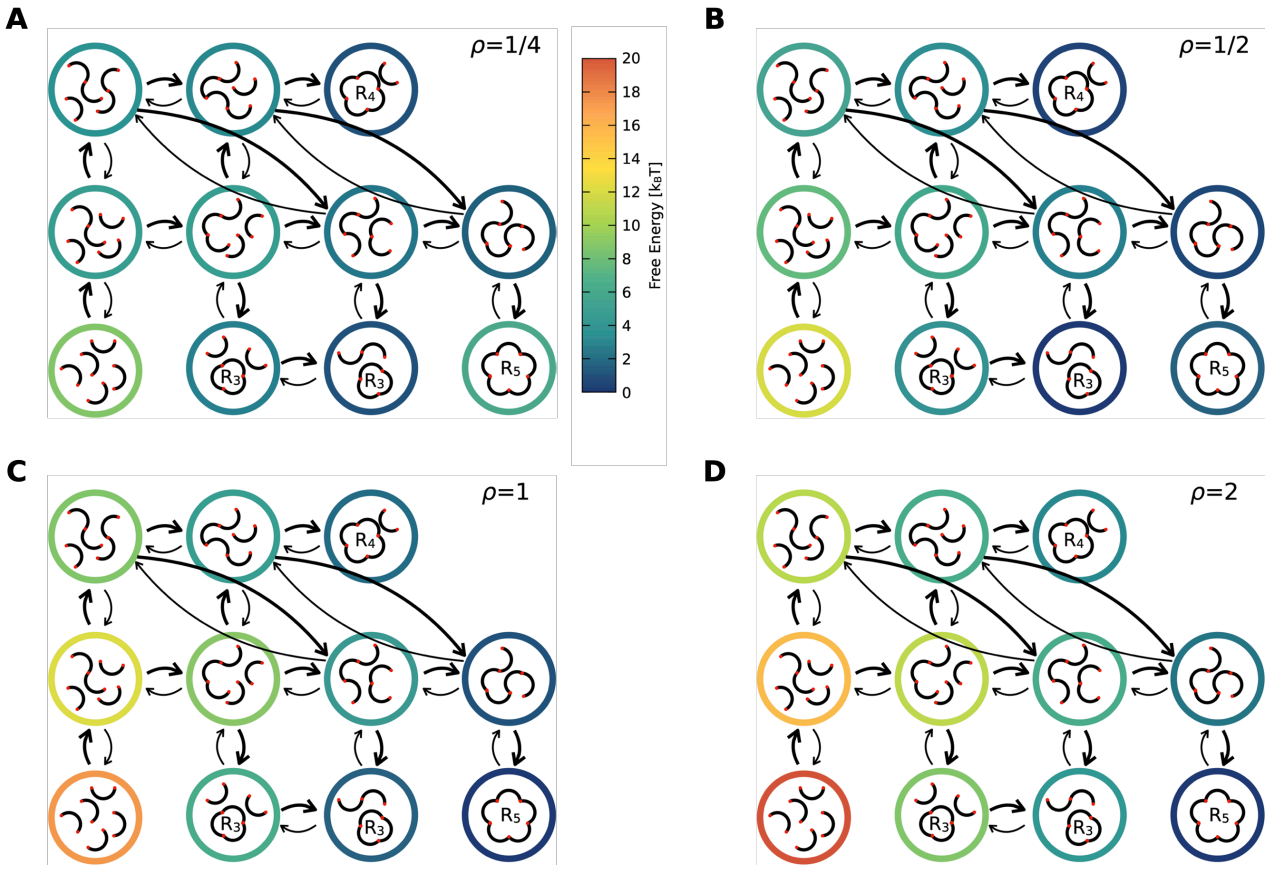

Fig. S 3 **Thermodynamics of a system composed of 5 building blocks at different building block concentrations.** The free energies associated with all the possible aggregation states (circles) are indicated by the color of each circle (in $k_{\mathrm{B}}T$, see color bar), while the arrows represent the transitions. Different panels correspond to different concentrations: (A) $\rho = 1/4$. (B) $\rho = 1/2$. (C) $\rho = 1$. (D) $\rho = 2$.

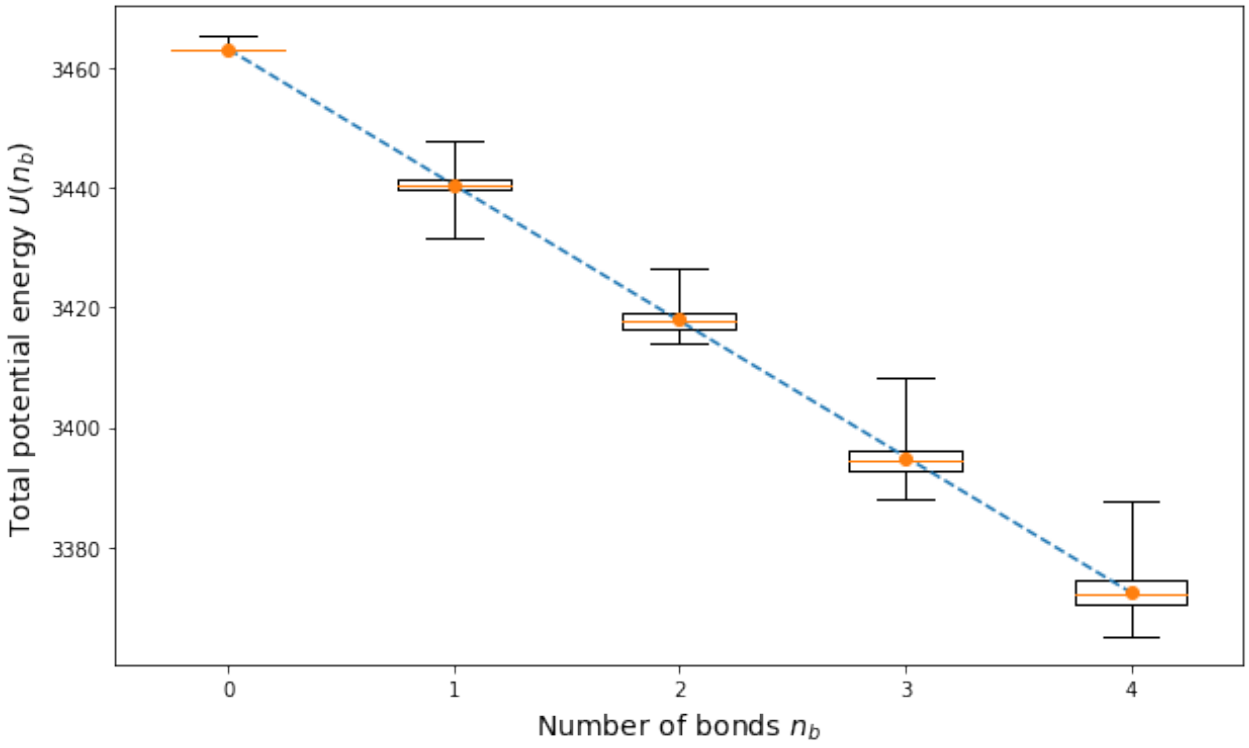

Fig. S 4 **Binding energy**. Box and whisker plot of the potential energy for increasing number of bonds (dot, average; center line, median; box limit, upper and lower quartile; whisker, min and max). The offset at $n_\mathrm{b} = 0$ is due to the computational model of the building block. The dotted blue line is a linear regression, which leads to a value of the enthalpic gain per bond equal to $\varepsilon_\mathrm{b} = -22.69\, k_\mathrm{B} T$.

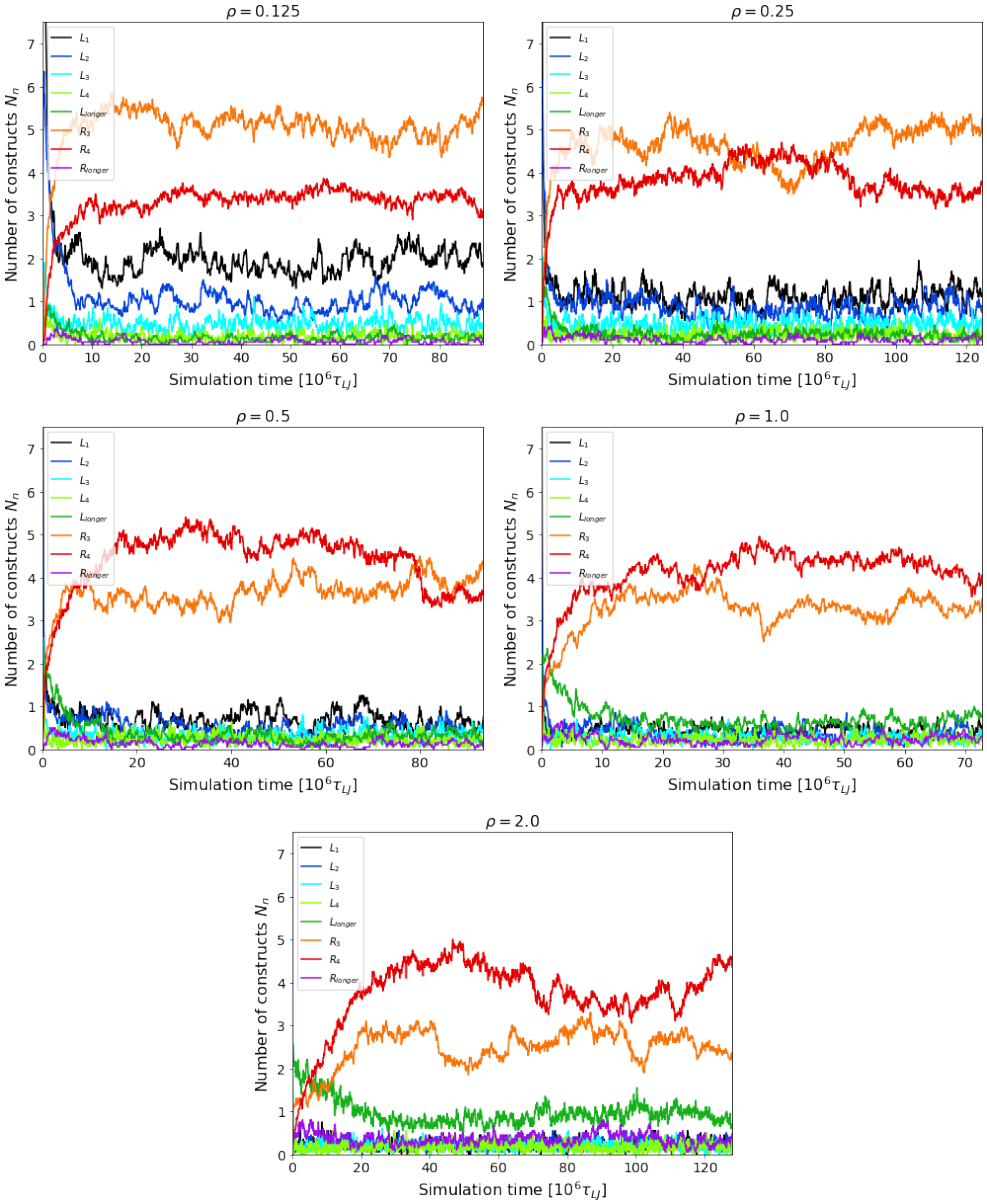

Fig. S 5 **Time evolution** of the average abundance of different species self-assembling from an initially random arrangement of $N = 36$ isolated building blocks at different concentrations ρ. Each trace represents the abundance averaged over 20 independent system realizations. As in the main text, linear species are labeled as L_n, while self-limited ones as R_n; n is the number of building blocks composing the construct. One can see that the self-limited assemblies R_3 and R_4 are always dominant over all the other species, and their relative concentrations can be tuned by varying the initial system density ρ.

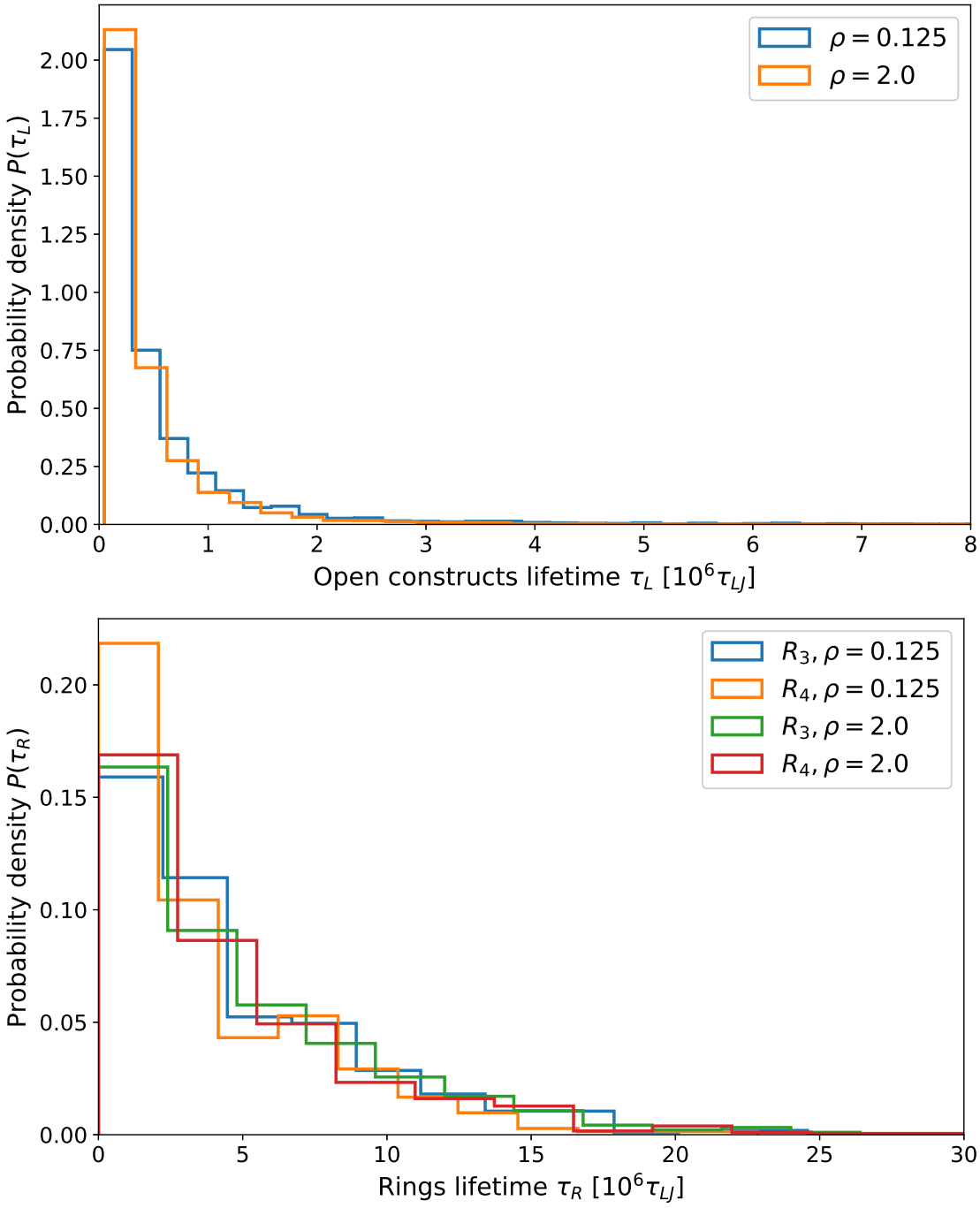

Fig. S 6 **Lifetimes distributions for different species. Upper panel**: distribution of open species' lifetimes at low and high system density. The distributions are indistinguishable and can be fitted with a decaying exponential. The average lifetime is around $0.5 \cdot 10^6 \tau_{\mathrm{LJ}}$. **Lower panel**: distribution of R_3 and R_4 lifetimes at low and high system density. In this case, too, the distributions are very similar and can be fitted with a decaying exponential. The average lifetime is around $4.5 \cdot 10^6 \tau_{\mathrm{LJ}}$.

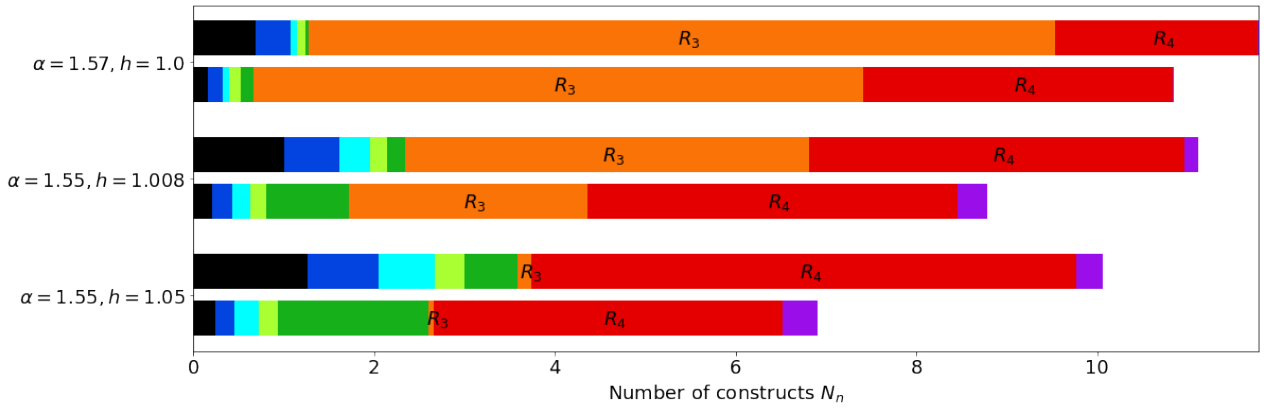

Fig. S 7 **Abundances N_n of the different species for three different building block geometries**, defined by the helix angle α and pitch h. For every geometry, the abundances at two system densities ($\rho = 0.25$, upper bars, and $\rho = 2.0$, lower bars) are shown. As can be seen, while the geometry used throughout all the work (central one in the figure, $\alpha = 1.55$ and $h = 1.008$) allows for the density-driven tunability, these data suggest that this property is lost for other, although similar, choices of the helix parameters, for which one of the two high-order assemblies is always the dominant one.