Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information

for

Improving the Productivity of Monodisperse Polyhedral Cages by Rational Design of Kinetic Self-assembly Pathways

Xiangze Zeng [a], Zhan-Wei Li *[b], Xiaoyan Zheng [a], Lizhe Zhu [a], Zhao-Yan Sun [b], Zhong-Yuan Lu [c], Xuhui Huang [a*]

[a] Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, E-mail: xuhuihuang@ust.hk

[b] State Key Laboratory of Polymer Physics and Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China, E-mail: zwli@ciac.ac.cn

[c] State Key Laboratory of Supramolecular Structure and Materials, Institute of Theoretical Chemistry, Jilin University, Changchun, China

1. Simulation details

In this work, we first studied the dodecahedral cage formation by the self-assembly of 3-patch particles as shown in Figure 1. The angles between the 3 patches are $120°, 120° \; and \; 108°$ to mimic the non-polar sp2 carbon atoms in fullerene. The half-opening angle of the patch is $50°$. α_{ij}^{R} and α_{ij}^{A} are 396 and 110 respectively, corresponding to approximately 4.0 MPa for the elastic modulus and 3.0 kBT for the adhesion energy if the particle size is 10 nm, respectively [1]. ν equals to 0.5. We refer this system as the "original system". To achieve a larger amount of dodecahedral conformations, we also changed all the angles between patches to $108°$. We refer this new system as "designed system". All the simulations were performed by the GALAMOST package in the NVT ensemble [2]. At the beginning, 960 patchy particles were solvated in the simulation box with size of $40 \times 40 \times 40$. Explicit coarse-grained solvent model is used in this study. The concentration of patchy particle was set to be 1.5%. The solute-solvent and solute-solute interactions are describe by the repulsive force in Equation 1 in the main text. The Velocity-Verlet algorithm [3] was used to integrate the equation of motions. The time step is $\delta t = 0.002\tau$ and the saving interval for each frame is 1000 steps. The temperature was coupled by Nose-Hoover thermostat [4]. We run 60 parallel trajectories with different starting velocities with the length of 2×10^6 steps for both of the original and designed systems.

2. Order parameters for state decomposition

In this study, we chose the size of the aggregate, number of pentagonal rings and number of hexagonal rings in the aggregate as order parameters to perform the state decomposition. The aggregate size is used to monitor the growth of the aggregates. We choose number of pentagonal and hexagonal rings as order parameters as they are two characteristic components of the closed fullerene-like cage formed by the patchy particles. As shown in Figure S1 (a) and (b), all extracted aggregate conformations are projected onto these three defined order parameters. Time evolution of these three order parameters of one simulation trajectory are presented in Figure S1 (c) and (d). More details of these three order parameters could be found in Section 3 & 4.

3. Extracting the aggregate conformations

We adopt the same method as described in ref [5] to identify aggregates. In particular, two patchy particles are considered as connected and belongs to the same aggregate if the distance between them is smaller than the distance of the first solvation shell (0.75) of their radial distribution functions (see Figure S2). In total, we extracted 10,945,409 and 13,150,914 aggregate conformations for the original and designed system, respectively.

4. Calculating the number of trigonal, tetragonal, pentagonal and hexagonal rings

The number of trigonal, tetragonal, pentagonal and hexagonal rings are calculated based on the "bonded" interaction between inter-particle patches. In this study, we consider that there is a "bond" between two inter-particle patches if there is an attractive interaction between these two inter-particle patches according to Equation 2 in the main text. If one patch does not have interaction with any other patches, we denote this patch as "un-bonded" patch. By this way, we could identify all the "bonds" between inter-particle patches in the system. Then one aggregate could be considered as a graph in which the node represents the patchy particle, the edge represents the "bond". And the rings in the aggregate could be regarded as the closed chordless cycles in the graph. The chordless cycle is a cycle in which two nodes are not connected by an edge not belong to the cycle. In this way, a closed cycle formed by the connection of two smaller closed cycles will not be counted as a new ring. By applying a greedy algorithm, we detected all the rings.

The probability of "un-bonded" patches (P_{unbond}) for each aggregate is calculated as the ratio of the number of "un-bonded" patches to the total number of patches in the aggregate.

5. State decomposition for the aggregate conformations

The aim of this study is to figure out the molecular mechanisms of the formation of the dodecahedral cage. Thus, we divided the aggregate conformations with size not larger than 34 into different conformational states based on the three order parameters defined in the previous section. The detailed protocols are listed below:

Step 1: We first picked up the aggregates with size smaller than 6 and the aggregates which are closed cages (with 12 pentagonal rings, N hexagonal rings and size equaling to 20+2N, N=0, 2, 3, 4, 5, 6, 7). These conformations are grouped as separated conformational states as shown in table 1.

Step 2: For the rest of the extracted aggregate conformations, those with the same number of pentagonal rings and same number of hexagonal rings are grouped into the same state.

Step 3: In each state obtained from step 2, the aggregate conformations are further divided into sub-states based on the following aggregate size range: $\{1-5\}, \{6-10\}, \{11-15\}, \{16-20\}, \{21-25\}, \{26-30\}$ and $\{31-34\}$.

These sub-states obtained from step 3 and the conformational states obtained from step 1 constitute all the conformational states. Finally, we obtained 663 and 1026 conformational states for the original and designed systems, respectively.

6. Identifying the self-assembly pathways based on the kinetic network model

To identify the self-assembly pathways for the formation of dodecahedral cage, we first constructed the kinetic network model based on the conformational states. The kinetics between different conformational states are characterized by the mass flow [5]. The lag time for calculate the mass flow between different conformational states is chosen to be 1000 time steps. Then we applied the modified Dijkstra's algorithm [6] on the net mass flow matrix. The initial state was chosen as the states containing aggregates with size smaller than 6 and without pentagonal or hexagonal rings (state S1 in Table 1). The final state was chosen as the state containing dodecahedral cages with size equaling to 20, 12 pentagonal rings and 0 hexagonal rings (state S2 in Table 1). In total, we identified 38 and 1860 pathways for the original and designed systems, respectively.

7. Testing the convergence of the kinetic analysis results

To test the convergence of the KNM analysis, 10, 20, 30, 40 and 50 trajectories are randomly chosen from all the 60 trajectories for the further analysis. We first examine the convergence of

the properties of individual conformational state in our KNMs. In particular, we chose five representative states at different stage of the self-assembly process as shown in Table 2. We show that state properties such as the resident mass flow (^{M_p}) and self-transition mass flow (T_{self}) calculated from different number of trajectories are all within the uncertainty (Figure S4). Here, the resident mass flow (^{M_p}) is defined as the total number of patchy particles (or accumulative mass of the aggregates) from a particular conformational state, and values of ^{M_p} relative to the averaged values of state A are reported for different number of trajectories in (Figure S4). The self-transition mass flow ($^{T_{self}}$) of a particular state is the diagonal term of the mass flow matrix corresponding to that state.

In addition to state properties, we have also examined the convergence of major self-assembly pathways (top pathways contribute to 60% of the total mass flow). As shown in Figure S5, the kinetic pathways are more difficult to converge than the state properties. In particular, top pathways computed from 10, 20, 30 and 40 trajectories display substantial difference from those calculated from 50 and 60 trajectory particularly at the initial stage of the self-assembly. Nevertheless, the kinetic pathways obtained from 50 and 60 trajectories are largely similar, indicating the convergence of our sampling.

8. Calculating the asphericity parameter of aggregates (^{A_p})

The asphericity parameter is applied to describe the morphology of the aggregate and is defined as follows:

$$A_p = \frac{\left(\lambda_1^2 - \lambda_2^2\right)^2 + \left(\lambda_1^2 - \lambda_3^2\right)^2 + \left(\lambda_3^2 - \lambda_2^2\right)^2}{2\left(\lambda_1^2 + \lambda_2^2 + \lambda_3^2\right)^2} \quad (S1)$$

where $\lambda_1, \lambda_2, \lambda_3$ are the three eigenvalues of the gyration tensor A:

$$A = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix} \quad (S2)$$

where $S_{xy} = \frac{1}{N}\sum_i (x_i - x_c)(y_i - y_c)$, N is the aggregate size, namely the number of patchy particles in the aggregate, (x_i, y_i, z_i) is the ith particle's coordinate and (x_c, y_c, z_c) is the mass center of the aggregate. A_p ranges from 0 to 1, 0 for a perfect sphere, 1 for a rod.

9. Hexagonal rings mediate the formation of dodecahedral cage structure

Although there is no hexagonal rings in the dodecahedral cage, we observed that hexagonal rings are formed in the intermediate states and then disappeared finally in the formation of the dodecahedral cage (Figure 2 (a)). We found that the conformational states containing hexagonal rings have a higher probability of the "bonded" patches, implying a higher stability of the structure. Further inspection of the simulation trajectories showed that the hexagonal rings is mainly formed by the coalescence of two tails connected to the same pentagonal ring as shown in Figure S6 (b), which is achieved by forming a new "bond" between particle 1 and particle 2. To form the dodecahedral structure, the hexagonal ring will convert to the pentagonal ring in the later stage as shown in Figure S6 (c), which is achieved by breaking the "bond" between particle 3 and particle 4 and forming a new "bond" between particle 3 and particle 5. Thus, we believe that the formation of hexagonal ring is an important intermediate state that mediates the formation of dodecahedral structure.

In the later stage, the hexagonal rings need to be converted to pentagonal rings in order to form the closed dodecahedral cage, which are only consisted of pentagon rings. This could be explained by the flexibility of the "bonds" formed between different patchy particles due to the relatively large size of the patches (see the patch opening angle defined in Figure 1). Though the angle between two patches within individual patchy patch is preciously 120° or 108°, the angles between two "bonds" formed by different patchy particles are not exactly 120° or 108°. Such flexibility allows the transition from the hexagonal ring to the pentagonal ring. Besides, to form the closed dodecahedral cage, an increase in the curvature of the structure also favors pentagonal rings.

10. The kinetics of the designed system differs from the original system

We found that the aggregates grow slower after we changed the angles as shown in Figure S8 (a). The growth of aggregates are controlled by two parameters: the adhesion rate characterizing the speed for forming larger aggregates and the dissociation rate characterizing the speed of the conversion from larger aggregates to smaller aggregates. We found the dissociation probabilities for the two systems are the same as shown in Figure S8 (b). Meanwhile, P_{unbond} of the designed system decreases more rapidly than the original system (Figure S8 (c)). These observations suggest that more aggregates with smaller size and smaller number of "un-bonded" patches are formed in the designed system than in the original system which in turn result in a smaller adhesion rate for the designed system. This indicates that we inhibit the formation of other aggregates larger than 20 to some extent by altering these 3 angles. We further identified the self-assembly pathways for the designed system as shown in Figure S9. Unlike the pathways in the original system, the aggregate size increasing stage is accompanied with the structure rearrangement process in the designed system.

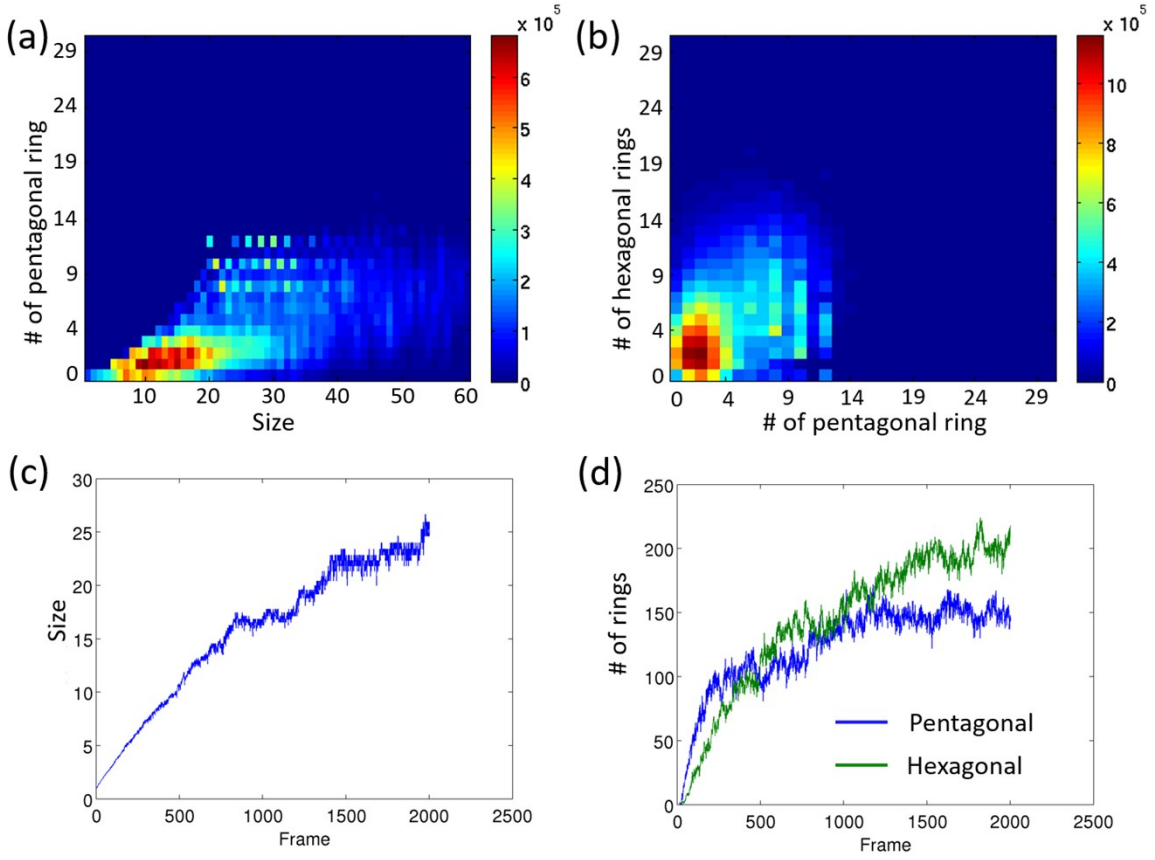

Figure S1. Projections of aggregate conformations onto the defined order parameters: (a) size and number of pentagonal ring; (b) size and number of hexagonal ring. The time evolution of (c) the mean aggregate size and (d) the number of the pentagonal and hexagonal rings in one simulation trajectory.

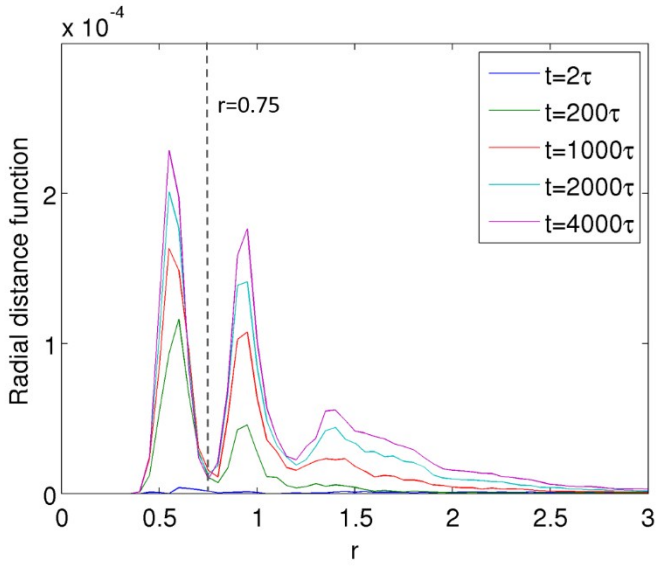

Figure S2. Radial distribution functions (RDF) of the patchy particles at different time points in the simulation. Each RDF plot is calculated based on one frame at the specified time point.

Table 1. Conformational states with specific size and number of pentagonal and hexagonal rings.

State label	Size	No. of pentagonal ring	No. of hexagonal ring
S1	0-5	0	0
S2	20	12	0
S3	24	12	2
S4	26	12	3
S5	28	12	4
S6	30	12	5
S7	32	12	6
S8	34	12	7

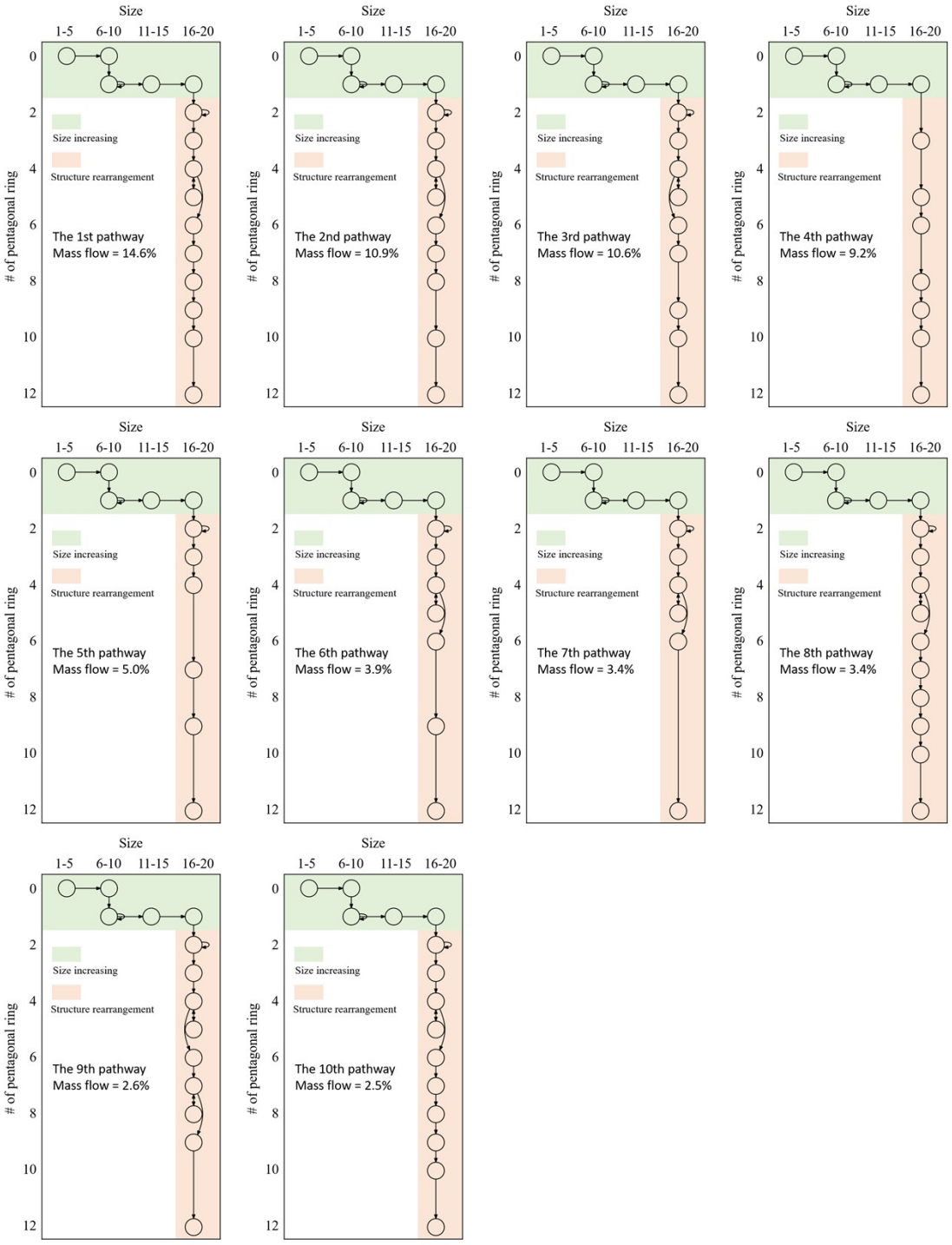

Figure S3. The top 10 pathways projected onto the size and the number of pentagonal rings. The corresponding mass flow of each pathway is labeled in the figure. Since the conformational state is defined based on the size, the number of pentagonal rings and the number hexagonal rings, we could see self-transition and reverse transition in the pathways projected onto the aggregate size and the number of pentagonal rings.

Table 2. Five representative states at different stage of the self-assembly process were chosen for the convergence test.

State label	Size	No. of pentagonal rings	No. of hexagonal rings
A	1-5	0	0
B	6-10	0	0
C	6-10	1	0
D	11-15	1	0
E	16-20	1	0

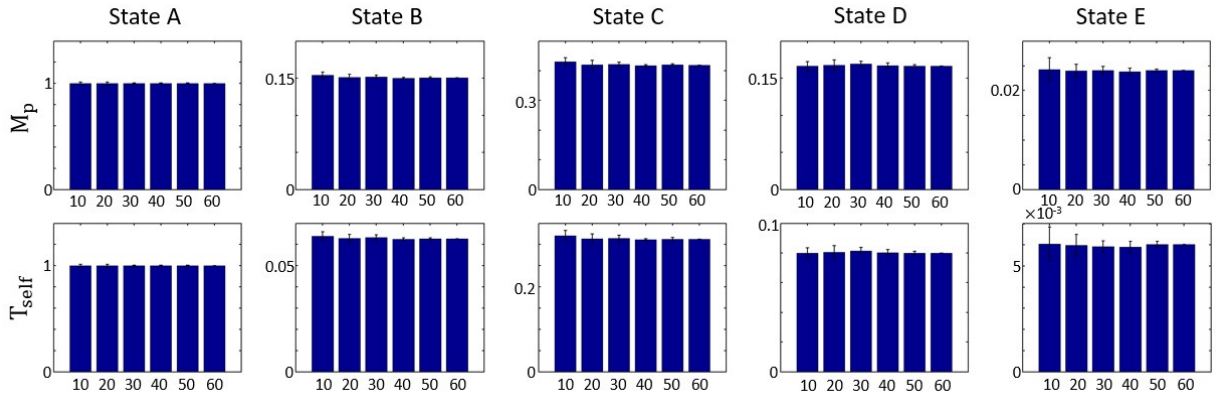

Figure S4. The relative mass of patchy particles (M_p) in the five conformational states and self-transition mass flow (T_{self}) of these five states. We calculated M_p and T_{self} by using different number of trajectories, namely 10, 20, 30, 40, 50 and 60 trajectories. The error bars were estimated by the bootstrapping method. Specifically, we bootstrapped 10 times, and at each time 10, 20, 30, 40 and 50 non-repeating trajectories were randomly chosen, respectively. The error bar for the data calculated from 60 trajectories is 0 as the total number of trajectories is 60. The values of M_p and T_{self} of State A were set to 1 and those of other states were scaled accordingly.

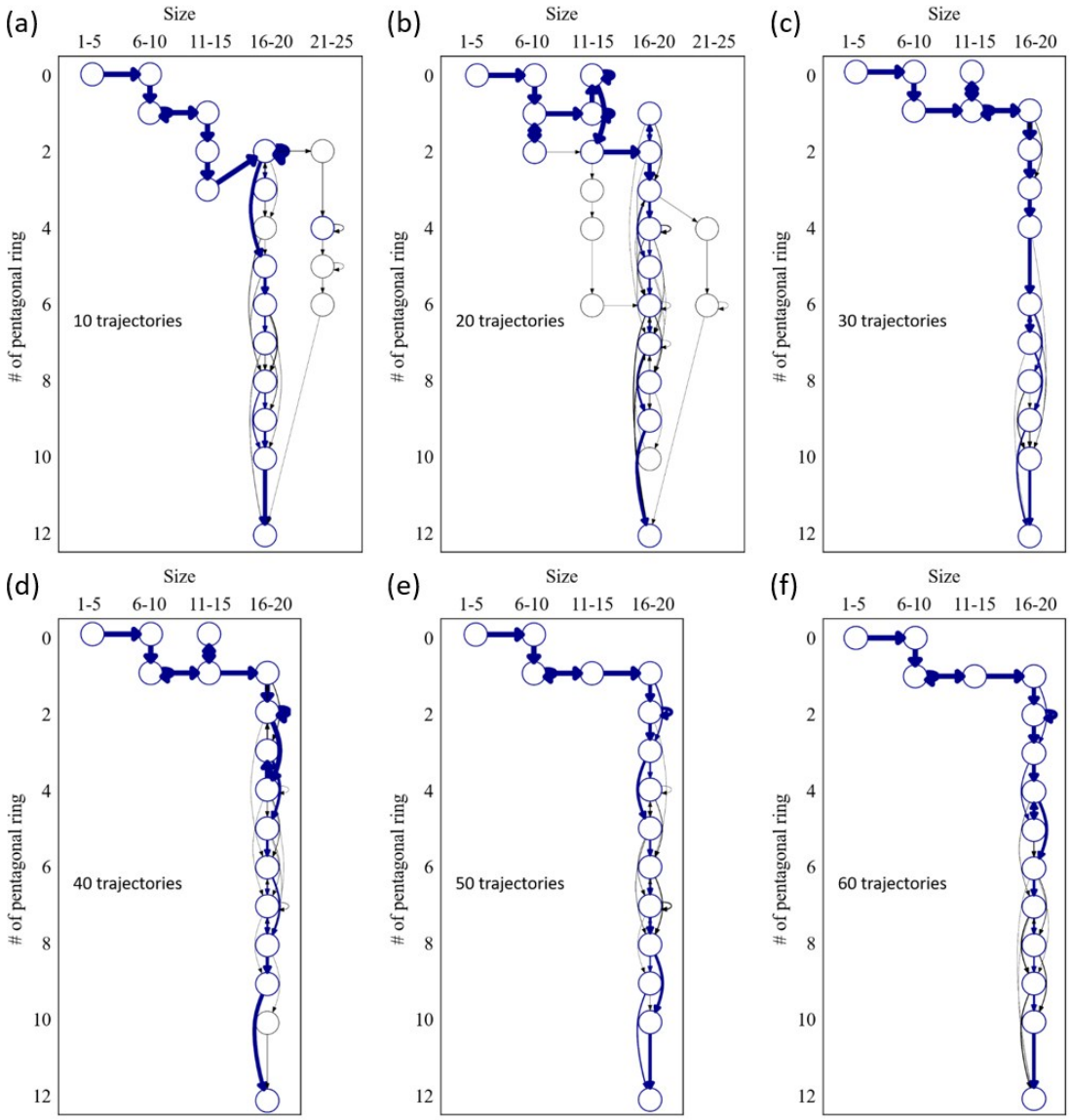

Figure S5. Top pathways conveying about 60% of the total mass flow are identified from the different KNMs constructed based on different number of trajectories.

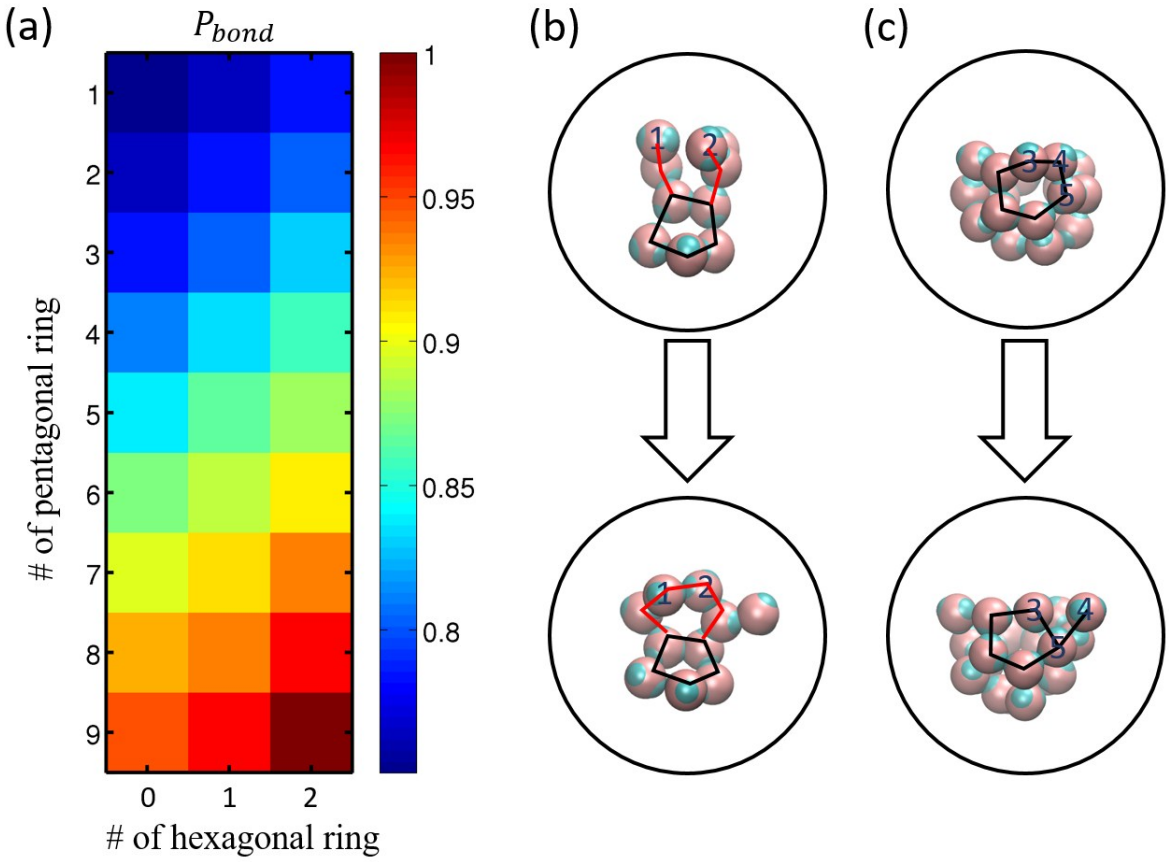

Figure S6. (a) Probability of "bonded" patches ($P_{bond} = 1 - P_{unbond}$) in the aggregate of the states projected onto the number of pentagonal rings and hexagonal rings. Conformations extracted from the simulation trajectory show: (b) the formation of a hexagonal ring and (c) the conversion from a hexagonal ring to a pentagonal ring.

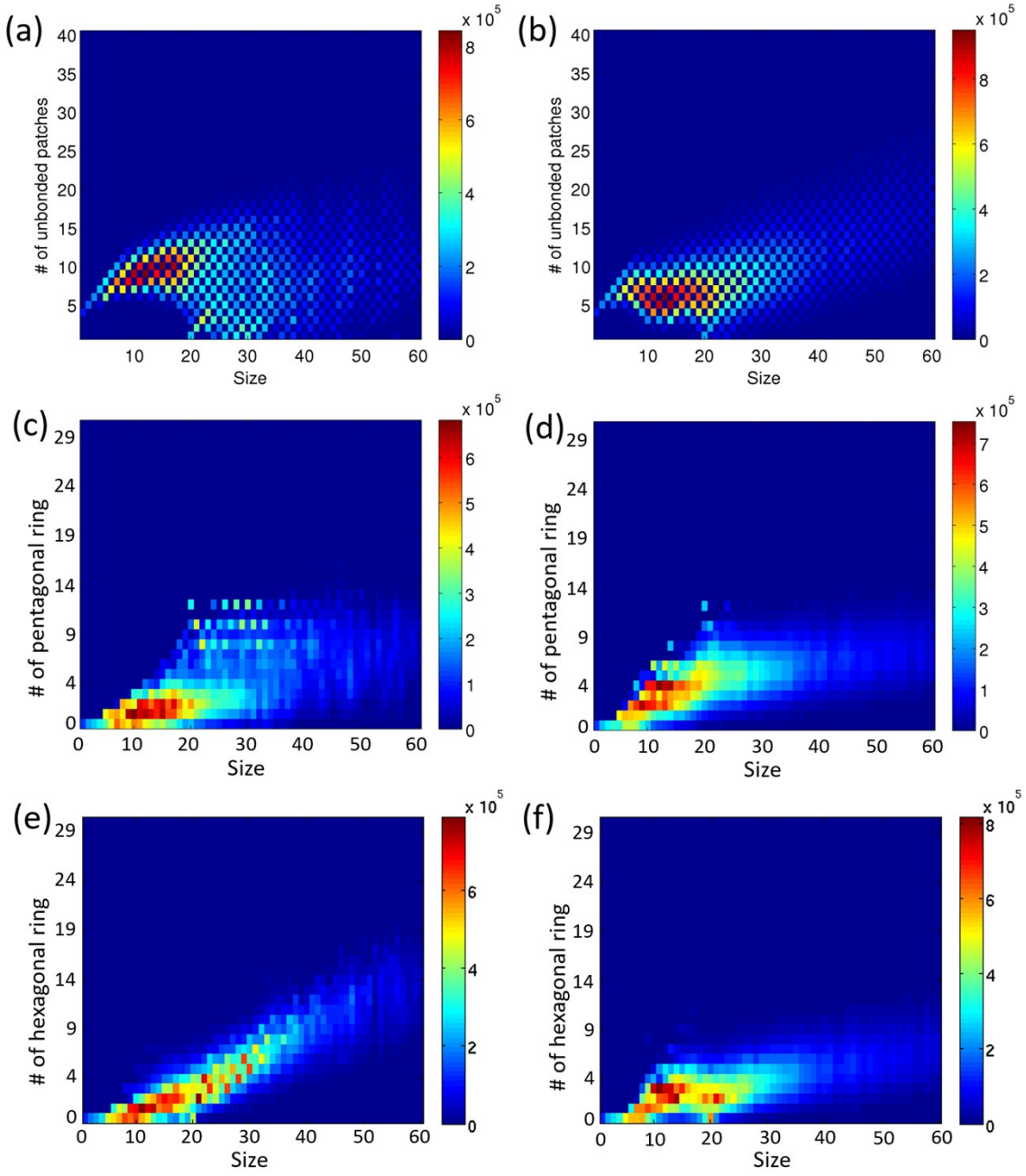

Figure S7. Projection of the aggregates onto the size and the number of "unbonded" patches, the number of pentagonal rings, the number of the hexagonal rings for the original system (a)(c)(e) and designed system (b)(d)(f), respectively.

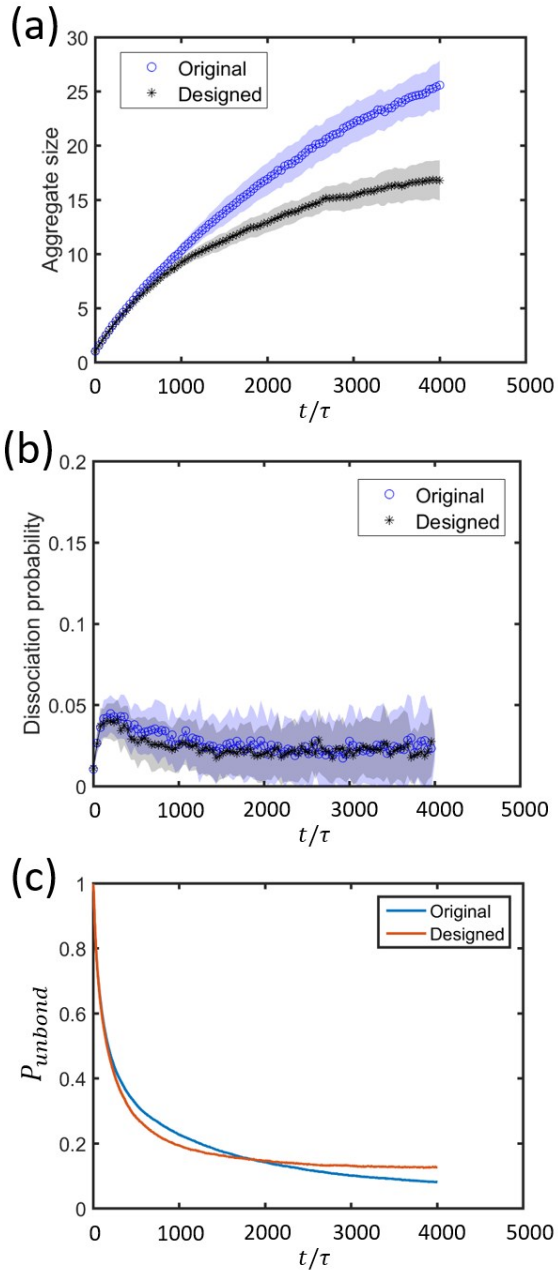

Figure S8. Time evolution of (a) mean aggregate size, (b) the dissociation probability, (c) the probability of un-bonded the patches.

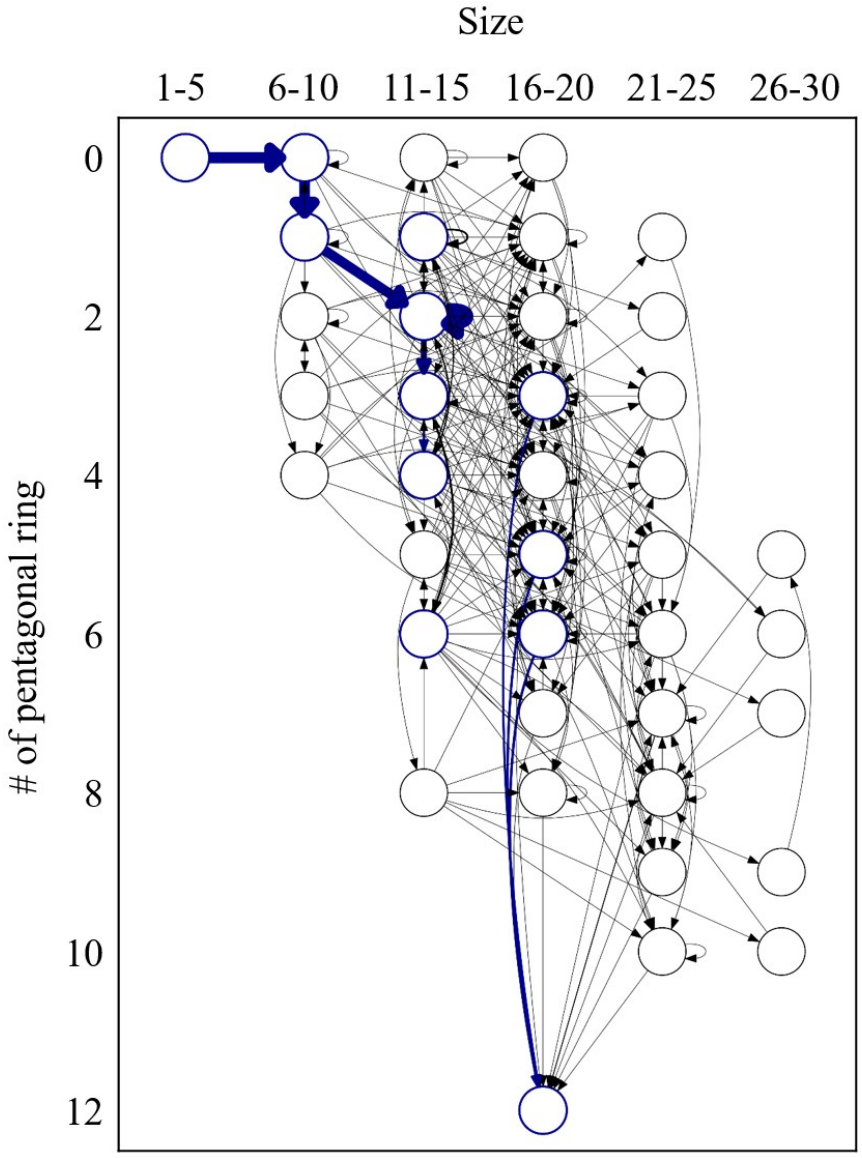

Figure S9. Top 600 pathways for the designed system projected onto size and the number of pentagonal rings. These top 600 pathways convey 62% of the mass flow of the system. The line width of the path segment connecting two states is proportional to the total flux in the top 600 paths that passes through these two states. Line width corresponding to flux less than 5% is set to be the line width corresponding to flux=5%. The path segment is colored blue if the corresponding flux is larger than 10%.

References

1. Li, Z.-W., et al., *Supracolloidal fullerene-like cages: design principles and formation mechanisms.* Physical Chemistry Chemical Physics, 2016. **18**(47): p. 32534-32540.
2. Zhu, Y.L., et al., *GALAMOST: GPU-accelerated large-scale molecular simulation toolkit.* Journal of computational chemistry, 2013. **34**(25): p. 2197-2211.
3. Allen, M.P. and D.J. Tildesley, *Computer simulation of liquids*. 2017: Oxford university press.
4. Frenkel, D. and B. Smit, *Understanding molecular simulation: from algorithms to applications*. Vol. 1. 2001: Academic press.
5. Zeng, X., et al., *Elucidating dominant pathways of the nano-particle self-assembly process.* Physical Chemistry Chemical Physics, 2016. **18**(34): p. 23494-23499.
6. Gupta, A., et al. *Free network measurement for adaptive virtualized distributed computing*. in *Proceedings 20th IEEE International Parallel & Distributed Processing Symposium*. 2006.