

Supplementary Information for Elucidating Dominant Pathways of the Nano-particle Self-Assembly Process

Xiangze Zeng,^a Bin Li,^{b,‡} Qin Qiao,^a Lizhe Zhu,^a Zhong-Yuan Lu^{*b} and Xuhui Huang^{*a,c}

1 Extracting the aggregate conformations

To judge whether or not the copolymers belong to the same aggregate, we adopt the concept of “contact pair”¹. We define the contact pair as a pair of inter-copolymer bead within a distance shorter than the cutoff of the non-bonded interaction r_{cut} . Two copolymers are considered connected and belong to the same aggregate if they contain at least 50 such contact pairs. Here we identified the contact pairs of copolymers by using the Quick Find algorithm². Finally, 115,786 conformations of aggregate were extracted from the Brownian Dynamics simulation dataset, which contains 25 simulations each with the length of 2.0×10^8 steps. With a save interval of 2.0×10^8 steps, we collected 25,000 simulation snapshots in total.

2 Order parameters to describe the self-assembly process

In this study we chose morphology and size to describe the evolution of aggregates during the self-assembly process. Among these two order parameters, the size of aggregate is a natural choice to monitor the growth of the self-assembled structures. In addition to size, different morphologies (shape) of aggregates, such as spherical micelle and rod during the copolymers self-assembly have also been widely reported in both experimental and computational studies³⁻⁹. In these studies, the asphericity parameter has often been adopted to distinguish different morphologies of aggregates¹⁰⁻¹⁵. These two order parameters can sufficiently describe various aggregate conformations along the self-assembly process. As shown in Fig. S1(b), the projections of the aggregate conformations onto the two order parameters span a wide range, indicating that the aggregate growth follows pathways consisting of various states of different shapes. Fig. S1(a) illustrates several typical aggregate structures: spheres, disks, worm-like rod, and lollipop-like rod. These conformations indicate highly diverse structures of intermediate states along the self-assembly pathways. Since aggregate with size of 150 is large enough to form vesicle as shown in Fig. S1(a), we focused our study on the aggregation process below size of 150.

3 Identifying metastable conformational states during the self-assembly process

The dynamics of the copolymer self-assembly process are complex and occur at a mixture of timescales. On one hand, during self-assembly, two or more small aggregates can diffuse to collapse and form a larger one. On the other hand, individual aggregate itself may also undergo conformational re-arrangements from the rod to spherical shape. As shown in Fig. 2(a), we found that the latter process (the individual aggregate shape re-arrangement) occurs at a much faster timescale than the former one (different aggregates diffuse to collapse) when the aggregate size is sufficiently large. In order to address this complex dynamics problem, we performed two independent state decompositions based on aggregate size and morphology (A_p) respectively, and subsequently combined them to produce the final set of kinetically metastable states.

Based on the aggregate size, we applied a typical splitting-and-lumping protocol to perform the state decomposition^{16,17}. At first, the k-centers clustering was performed to split all aggregate conformations with the size up to 150 (73,816 conformations in total) into 20 microstates according to their size (see Fig. S2(a)). This will allow us to investigate how star-like block copolymers self-assemble to form vesicles, as aggregates with the size of 150 are already large enough to form distinct vesicle structures (see Fig. S1(a)). Next, we performed the kinetic clustering using the Perron Cluster Cluster Analysis (PCCA) method¹⁸ to lump these 20 microstates into 7 metastable macrostates. Rather than using the transition probability matrix, we here adopted the mass flow matrix (see Equation (1)) upon symmetrization to characterize the transitions between microstates. Along the progress of self-assembly, the number of aggregates in the simulation box decreases as their sizes increase, this causes transitions across a large size range when two big aggregates with similar size collapse. These long-range transitions may cause microstates with quite different sizes to be grouped into a single macrostate, which prevents us to obtain clear visualizations of the self-assembly pathways. Therefore, we only take into account transitions with the aggregate's size increase up to 20%. We note that this is only applied to the kinetic lumping step, and we still use the original un-modified mass flow matrix to identify the dominant self-assembly pathways (see Section 4). We also show that the main

^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 Supramolecular Structure and Materials, Institute of Theoretical Chemistry, Jilin University, Changchun, China; E-mail: luzhy@jlu.edu.cn

^c Division of Biomedical Engineering, Center of Systems Biology and Human Health, Institute for Advance Study, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

‡ Present address: Theoretical Chemistry, Chemical Centre, Lund University, P.O. Box 124, S-221 00 Lund, Sweden

feature of the dominant self-assembly pathway, i.e. the growth of aggregate follows the “rod-to-disk-to-sphere” cycle, does not change when the number of metastable macrostates is varied in the kinetic lumping (see Fig. S6).

For the aggregate morphology, we simply performed a k-means clustering to divide the aggregate conformations into 3 groups based on the value of A_p : 0-0.14; 0.14-0.46, and 0.46-1, which correspond to sphere, disk and rod shape respectively (see Fig. S2(c) and S3(b)). These three distinct shapes we identified are consistent with what have been reported in previous studies^{4,19}. When performing the clustering, we left out aggregates with size not larger than 5, because it is difficult to clearly define the morphology of aggregates at this small size. We note that the spheres could be further divided to spherical micelles, semi-vesicles and vesicles based on their aggregate size in our work. Spherical micelles are the aggregates that consist of a hydrophobic core surrounded by the hydrophilic part of the aggregate²⁰ as the insert in Fig. S4(a). Semi-vesicles are the aggregates that consist of a hydrophilic core, an intermediated hydrophobic layer and a hydrophilic layer²¹ as the insert in Fig. S4(b). Vesicles are hollow spheres and formed by the closure of the copolymer bilayers²⁰ as the insert in Fig. S4(c). These differences between them could be characterized by the radial density distribution of the hydrophobic and hydrophilic beads as a function of the distance to the mass center as shown in Fig. S4(a)-(c). For spherical micelles, there are no A beads within 1σ of the aggregate center of mass (see Fig. S4(a)). For semi-vesicles, there are no B beads within 1σ of the aggregate center of mass (see Fig. S4(b)). For vesicles, there are neither A beads nor B beads within 1σ of the aggregate center of mass (Fig. S4(c)). We found the identified metastable states based on the aggregate size and the kinetics are sufficient to distinguish these structures as showed by the probability of finding certain number of A beads and B beads within 1σ of the aggregate center of mass in Figure S4(d)-(f). In addition, we calculated the radius of gyration of the aggregates. The mean values of the radius of gyration of the semi-vesicles with size from 41 to 76 and the vesicles with size from 120 to 150 are 8.4σ (about 10nm) and 10.9σ (about 13nm), respectively.

At last, we combined the results of the two above-mentioned clustering as two independent sets, and obtained 19 states in our final model (see Fig. S2(d)). We performed the state decomposition on the simulations with poor solvent condition (with $\alpha = 0.6$). In order to examine how dominant self-assembly pathways are altered with the change of α (from 0.6 to 0.4, from poor solvent condition to good solvent condition), we need consistent state decomposition between the two systems. Therefore, we simply assigned the aggregate conformations from the good solvent condition system based on the state definition obtained from the poor solvent condition system.

4 Identifying the dominant self-assembly pathways

In this study, nearly all the transitions we observed are unidirectional from small to large aggregate size. In this case, even though the individual Brownian dynamics simulations satisfy the detailed balanced, the transitions corresponding to the aggregate dissociation are difficult to be observed within the timescale we simulate, because the copolymer self-assembly is a strong energetically downhill process to favor the aggregated states. As a result, the Transition Path Theory (TPT)^{22,23} is not directly applicable to our dataset as it is formulated under the condition of equilibrium sampling. To avoid this issue, we have identified the self-assembly pathways based on the flow of mass rather than the equilibrium flux in TPT. In particular, we define the mass flow from state i to j after a lag time of among all the simulation trajectories:

$$u_{ij} = \sum_{trj} \sum_k \sum_{n=1}^{N-1} m S_k \delta_{x(k,n\Delta t)}^i \delta_{x(k',(n+1)\Delta t)}^j \quad (1)$$

where trj represents the trajectory id, k represents the aggregate id in one trajectory (one aggregate is assigned with the same id before it collides with another one), k' is the aggregate id for aggregate k after time Δt , m is the mass of one copolymer, S_k is the size of the aggregate k , $\delta_{x(k,t)}^i$ is the state indicator function:

$$\delta_{x(k,t)}^i = \begin{cases} 1 & \text{if } x(k,t) = i \\ 0 & \text{if } x(k,t) \neq i \end{cases} \quad (2)$$

where $x(k,t)$ denotes to the state id that aggregate k belongs to at time t . The net mass flow u_{ij}^{\dagger} from state i to j is then defined as

$$u_{ij}^{\dagger} = \max(u_{ij} - u_{ji}, 0) \quad (3)$$

Using the net mass flow matrix, we can identify dominant pathways of the self-assembly process from initial state I to final state F (see Fig. S2(d)). To achieve this, we have converted the net mass flow matrix to a directed graph in which nodes and edges correspond to the states and pairwise mass flow respectively. In this case, identifying the dominant self-assembly pathway is equivalent to solving the widest path problem²⁴. Using the modified Dijkstra’s algorithm²⁴, we can identify the most probable pathway between the initial and final state containing the maximum net mass flow. To find the n th widest pathway, we adopted the same procedure applied on the net flux matrix in TPT implemented in MSMBuild2²⁵: (1) Remove the bottleneck of the $(n-1)$ th major pathway; (2) Find n th major pathway by a relaxation process. By repeating these two steps, we can identify all the pathways.

5 Formation of vesicles via the semi-vesicles

We noted that the transition from semi-vesicle to vesicle in our study differs from that in the previous study^{21,26} where the semi-vesicle grows into larger vesicle purely by the diffusion of hydrophilic beads and solvent into the core without coalescence with other micelles

or semi-vesicles. In the present system, however, the semi-vesicle transforms into a vesicle by colliding with another aggregate and forming a bowl-like structure (Fig. S11). This implies that the vesicle formation process via the intermediate semi-vesicle structure in the present study is a mixture of diffusion and membrane closure mechanisms.

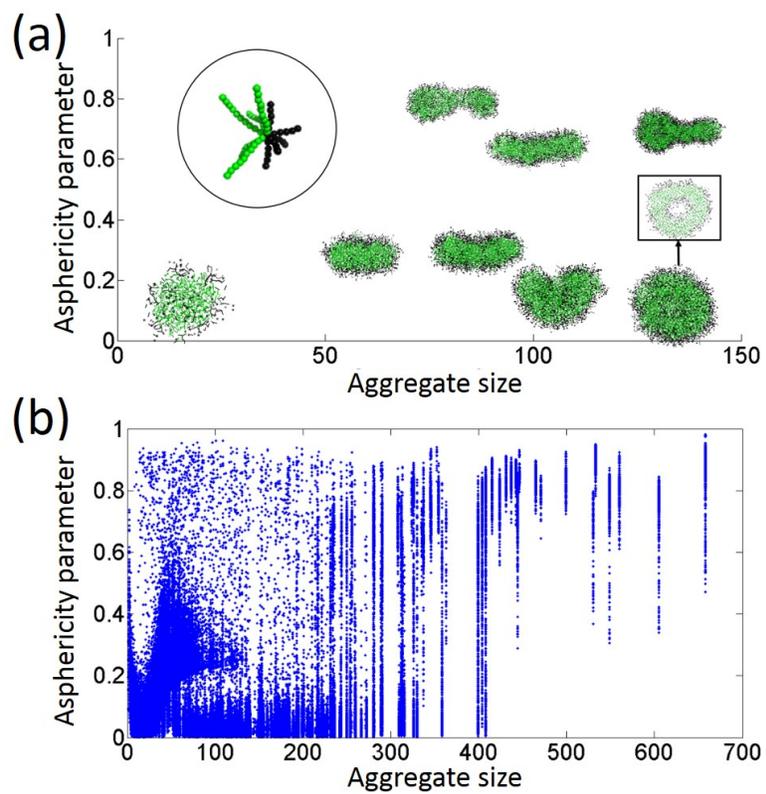


Fig. S1 (a): Typical aggregate structures with various size and shape are projected onto two order parameters: asphericity parameter (A_p) and aggregate size. The coarse-grained structure of one star-like block copolymer is shown in the inset. The hydrophilic beads (A beads) and hydrophobic beads (B beads) are shown in black and green respectively. The cross section of one sphere is shown in the rectangle. (b) Projections of all extracted aggregate conformations onto these two order parameters.

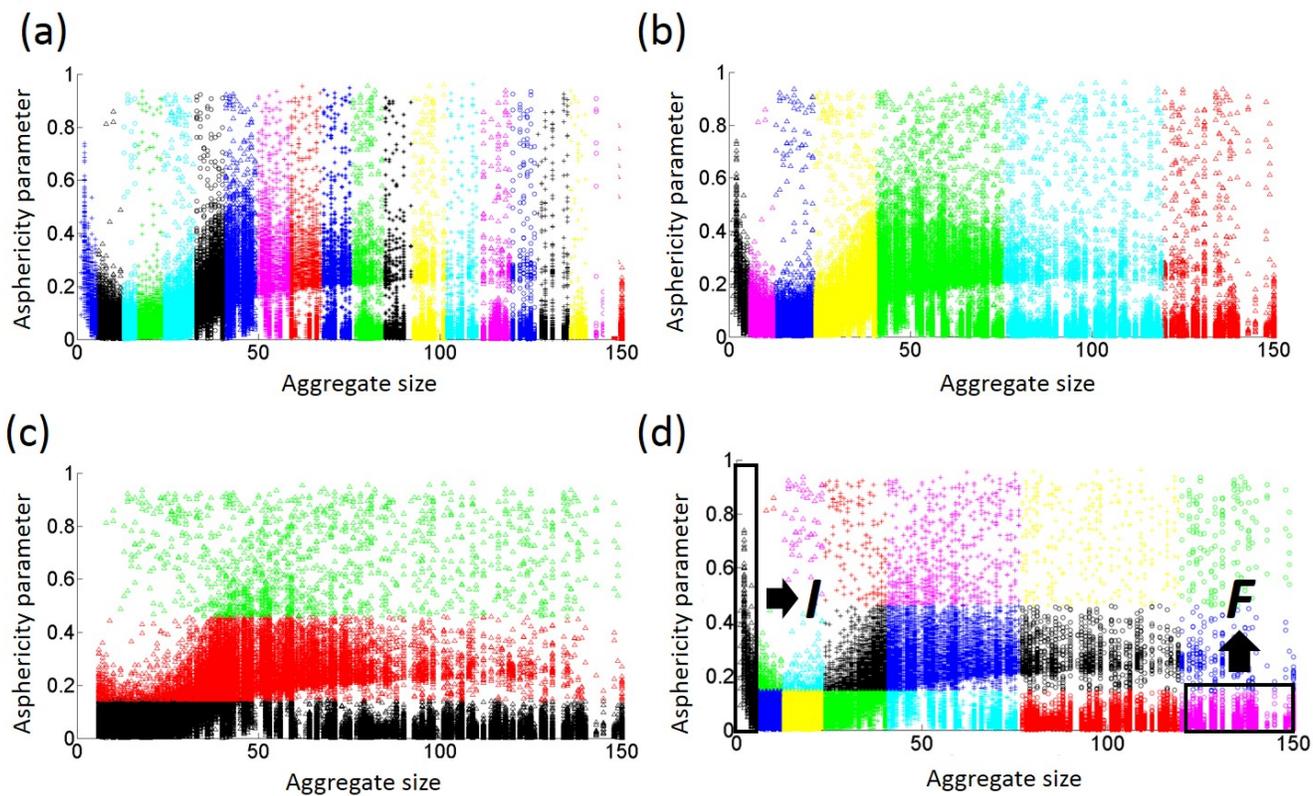


Fig. S2 (a) 20 microstates represented by 7 different colors were obtained by applying k-centers algorithm based on aggregate size. (b) 7 macrostates represented by 7 different colors were identified by PCCA method. (c) 3 morphological clusters represented by 3 different colors were obtained by applying k-means algorithm based on A_p . We note that the figure does not include the data of aggregate with size not larger than 5. (d) By combing clusters in (b) and (c), we obtained the final 19 metastable states represented by 7 different colors. State *I* and *F* were chosen as the initial and final state to identify dominant self-assembly pathways.

(a)

Cluster	Aggregate size	Size label
1	1 - 5	Small
2	6 - 12	
3	13 - 23	
4	24 - 40	
5	41 - 76	Middle
6	77 - 119	Large
7	120 - 150	

(b)

Cluster	Asphericity parameter	Shape
1	0.46 – 1.0	Rod
2	0.14 – 0.46	Disk
3	0.0 – 0.14	Sphere

Fig. S3 (a) 7 macrostates obtained based on aggregate size and the corresponding aggregate size range. (b) 3 morphological clusters obtained based on the asphericity parameter (A_p) and the corresponding A_p range.

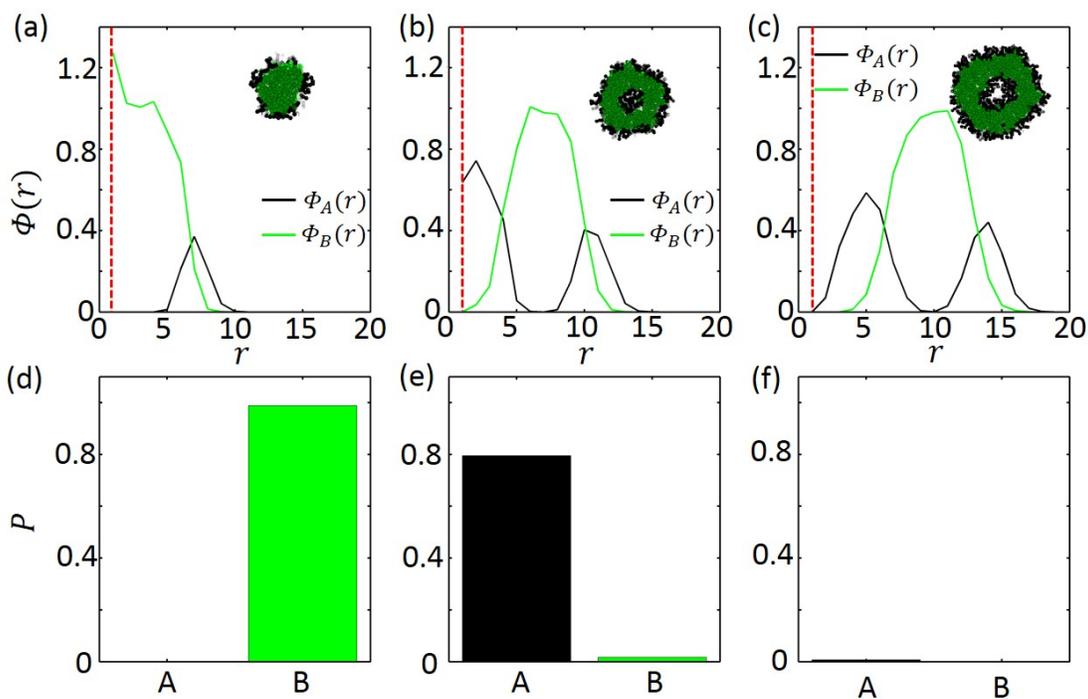


Fig. S4 Radial density distribution $\Phi(r)$ of the hydrophobic A beads and hydrophobic B beads as a function of the distance to the mass center (r) (a) in a spherical micelle, (b) semi-vesicle and (c) vesicle, respectively. The distribution is normalized such that $\int_0^R 4\pi r^2 \Phi_A(r) dr = N_A$ and $\int_0^R 4\pi r^2 \Phi_B(r) dr = N_B$. N_A and N_B are the total number of hydrophilic A beads and hydrophobic B beads, respectively, in the aggregate. The red dash line locates the position that is 1σ far away from the mass center. The cross sections of the aggregates are shown in the insets. The probability (P) of finding more than 7 hydrophilic A beads (black bar) and hydrophobic B beads (green bar) within the distance of σ from the mass center of the aggregates can be found in the (d) for spherical micelles with size from 13 to 23, (e) for semi-vesicles with size from 41 to 76 and (f) for vesicles with size from 120 to 150.

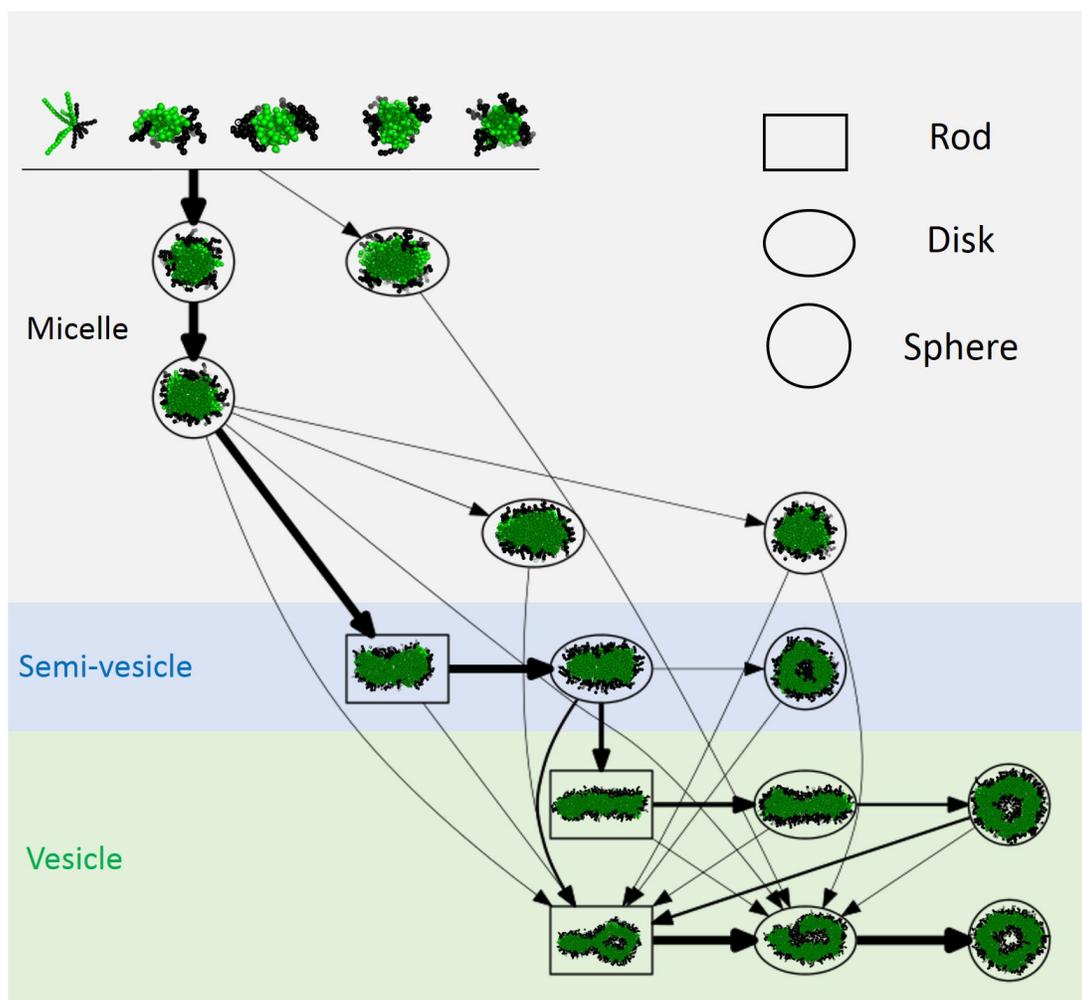


Fig. S5 All the identified self-assembly pathways in poor solvent condition. Line width is proportional to the total flux between the states connected by the line. Line width that corresponds to flux less than 10% is set to be the line width that corresponds to flux=10%. Three different background colors are used to distinguish different stages of aggregate growth as Figure 1.

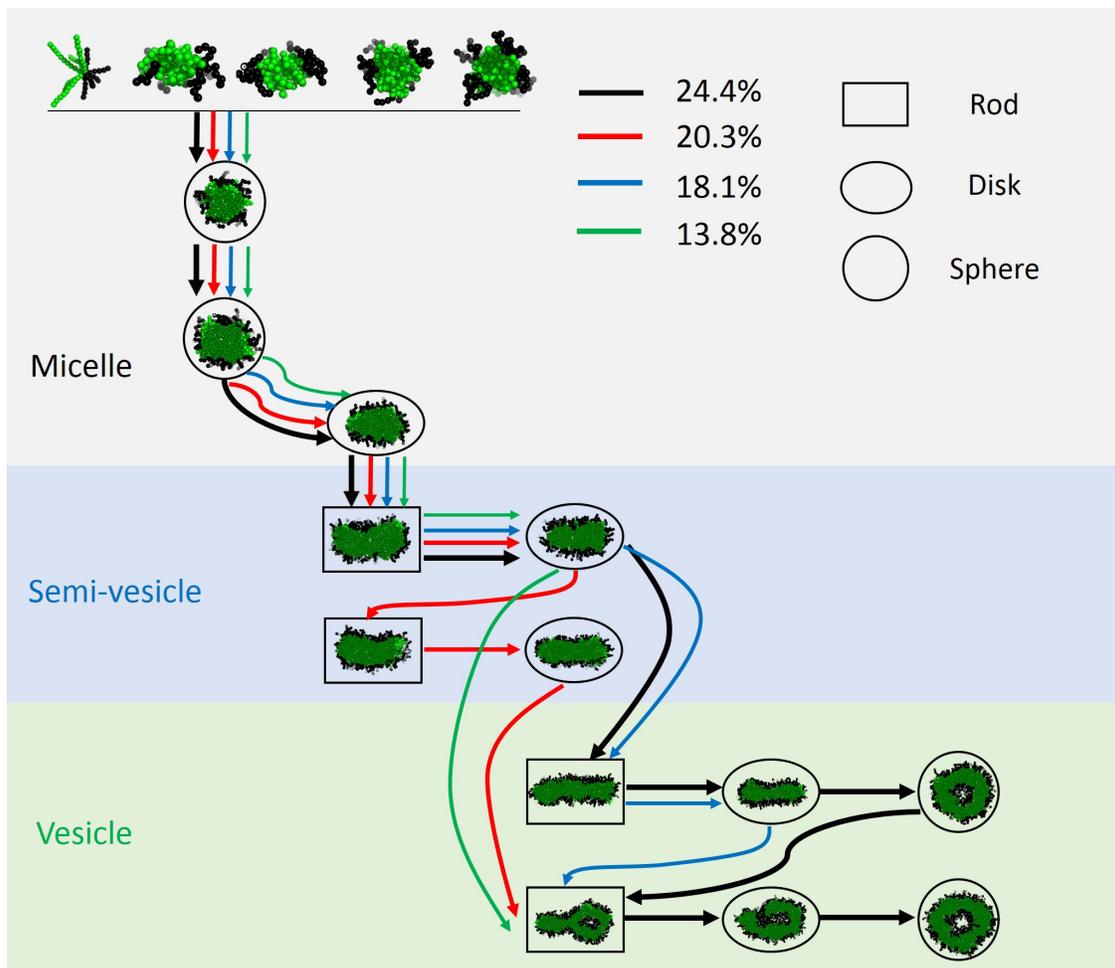


Fig. S6 Top 4 self-assembly pathways identified after we changed the number of metastable macrostates to 8 based on aggregate size.

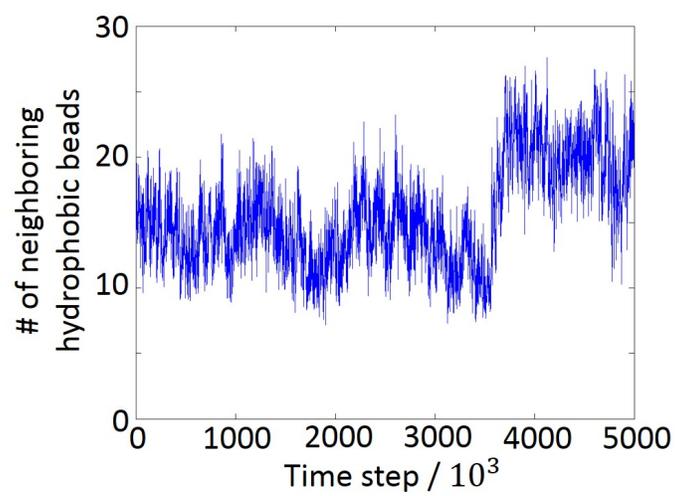


Fig. S7 Time evolution of the average number of hydrophobic beads around each hydrophilic bead within a distance of 3σ in one copolymer that diffuses into the core of the aggregate in the semi-vesicle stage. The starting point of x axis is same time with Figure 2(c) in the simulation.

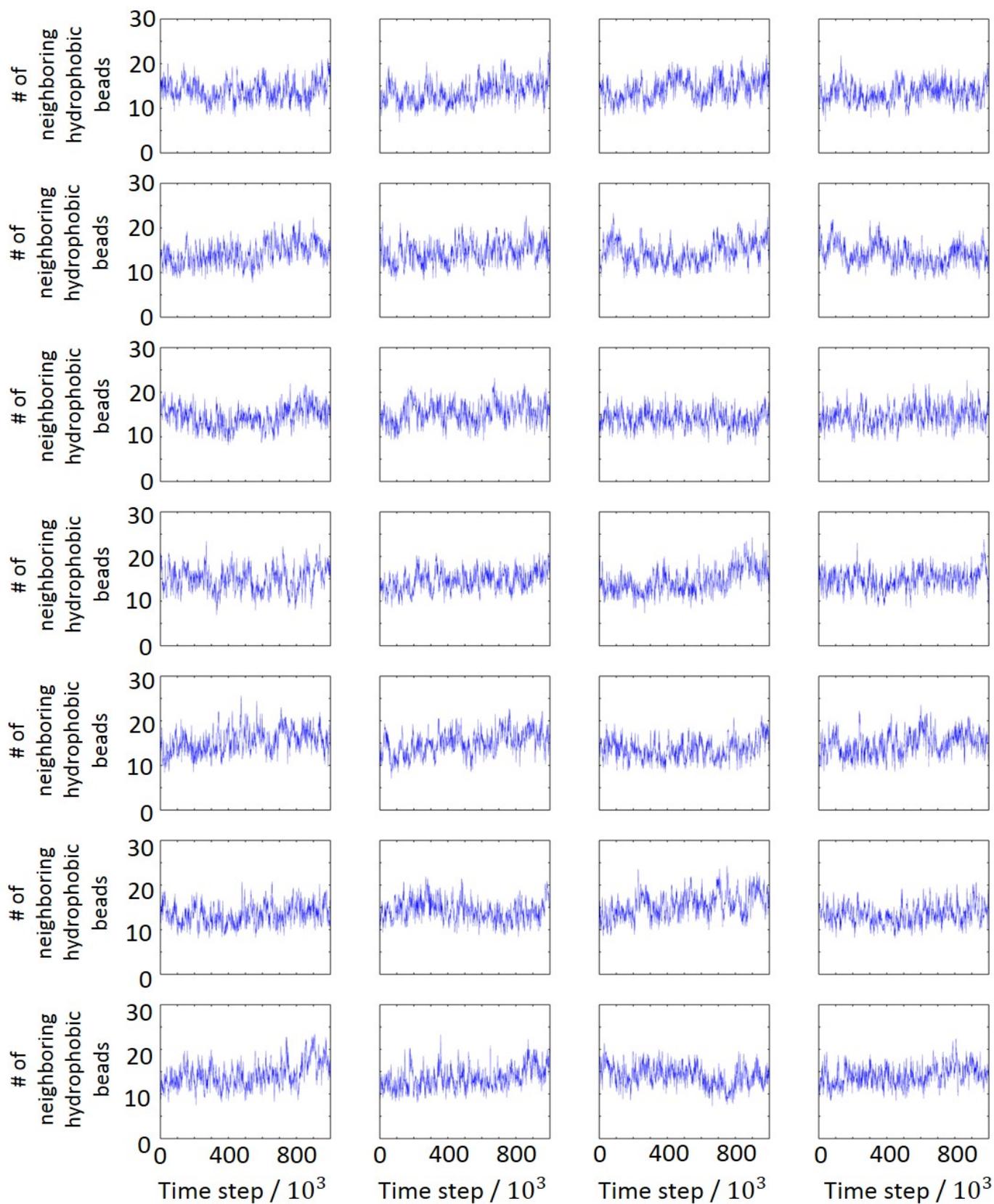


Fig. S8 Time evolution of the average number of hydrophobic beads around each hydrophilic bead within a distance of 3σ in one copolymer that forms the inner layer of the vesicle.

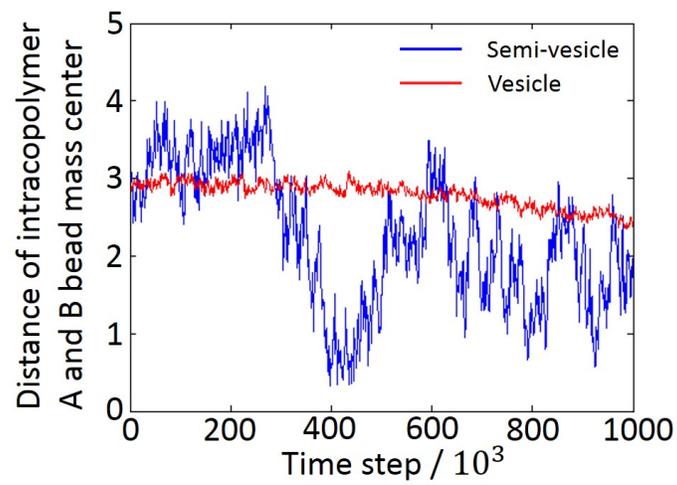


Fig. S9 Evolution of the distance between intra-copolymer A and B bead mass center during the formation of the semi-vesicle (blue line) and vesicle (red line).

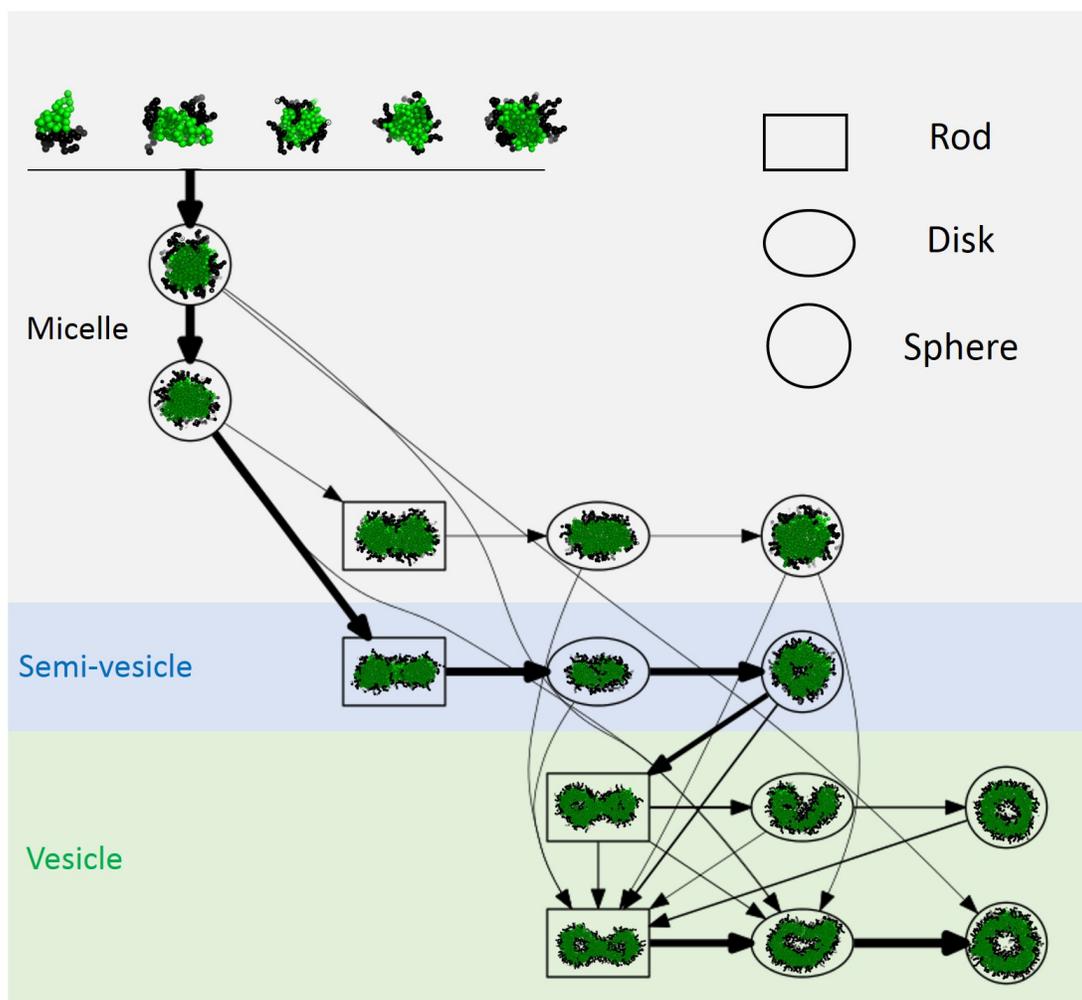


Fig. S10 All the identified self-assembly pathways in good solvent condition. Line width is proportional to the total flux between the states connected by the line. Line width that corresponds to flux less than 10% is set to be the line width that corresponds to flux=10%. Three different background colors are used to distinguish different stages of aggregate growth as Figure 1.

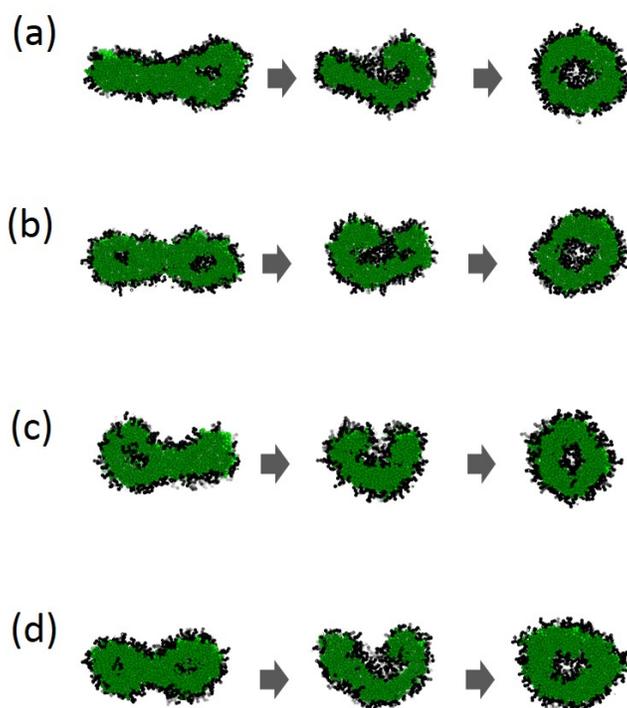


Fig. S11 The transformation of one semi-vesicle into vesicle after collapsing with another aggregate. (a) Collapse with one disk-like micelle in poor solvent condition. (b) Collapse with another semi-vesicle in poor solvent condition. (c) Collapse with one disk-like micelle in good solvent condition. (d) Collapse with another semi-vesicle in good solvent condition.

References

- 1 Y. I. Yang and Y. Q. Gao, *J. Phys. Chem. B*, 2014, **119**, 662–670.
- 2 R. Sedgewick, *Algorithms in C*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1990.
- 3 J. Zhu, Y. Jiang, H. Liang and W. Jiang, *J. Phys. Chem. B*, 2005, **109**, 8619–8625.
- 4 L. Zhang and A. Eisenberg, *J. Am. Chem. Soc.*, 1996, **118**, 3168–3181.
- 5 Y. Yu, L. Zhang and A. Eisenberg, *Macromolecules*, 1998, **31**, 1144–1154.
- 6 M. Han, M. Hong and E. Sim, *J. Chem. Phys.*, 2011, **134**, 204901.
- 7 A. Choucair and A. Eisenberg, *Eur. Phys. J. E*, 2003, **10**, 37–44.
- 8 M. R. Perrett and M. F. Hagan, *J. Chem. Phys.*, 2014, **140**, 214101.
- 9 Y. Han, H. Yu, H. Du and W. Jiang, *J. Am. Chem. Soc.*, 2010, **132**, 1144–1150.
- 10 J. Rudnick and G. Gaspari, *J. Phys. A. Math. Gen.*, 1986, **19**, L191.
- 11 G. Zifferer, *J. Chem. Phys.*, 1995, **102**, 3720–3726.
- 12 H. Noguchi and K. Yoshikawa, *J. Chem. Phys.*, 1998, **109**, 5070–5077.
- 13 T. Zhou and S. B. Chen, *Macromolecules*, 2005, **38**, 8554–8561.
- 14 C.-M. Lin, Y.-Z. Chen, Y.-J. Sheng and H.-K. Tsao, *React. Funct. Polym.*, 2009, **69**, 539–545.
- 15 J. Zhang, Z.-Y. Lu and Z.-Y. Sun, *Soft Matter*, 2011, **7**, 9944–9950.
- 16 J. H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noe, *J. Chem. Phys.*, 2011, **134**, year.
- 17 G. R. Bowman, X. Huang and V. S. Pande, *Methods*, 2009, **49**, 197–201.
- 18 P. Deuffhard, W. Huisinga, A. Fischer and C. Schütte, *Linear Algebra Appl.*, 2000, **315**, 39–59.
- 19 M. Xiao, G. Xia, R. Wang and D. Xie, *Soft Matter*, 2012, **8**, 7865–7874.
- 20 Y. Mai and A. Eisenberg, *Chem. Soc. Rev.*, 2012, **41**, 5969–5985.
- 21 X. He and F. Schmid, *Macromolecules*, 2006, **39**, 2654–2662.
- 22 W. E and E. Vanden-Eijnden, *J. Stat. Phys.*, 2006, **123**, 503–523.
- 23 P. Metzner, C. Schütte and E. Vanden-Eijnden, *Multiscale Model. Simul.*, 2009, **7**, 1192–1219.
- 24 A. Gupta, M. Zangrilli, A. I. Sundararaj, A. I. Huang, P. A. Dinda and B. B. Lowekamp, *Parallel Distrib. Process. Symp. 2006. IPDPS 2006. 20th Int.*, 2006, pp. 10 pp.–.

- 25 K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque and V. S. Pande, *J. Chem. Theory Comput.*, 2011, **7**, 3412–3419.
- 26 X. He and F. Schmid, *Phys. Rev. Lett.*, 2008, **100**, 137802.