Electronic Supplementary Material (ESI) for Soft Matter.

Supplementary Information

Self-assembly prediction of architecture-controlled bottlebrush copolymers in solution using graph convolutional network

Wooseop Hwang, [a, †] Sangwoo Kwon, [b, †] Won Bo Lee,*[,b] and YongJoo Kim*[,a]

[a]Department of Materials Science and Engineering, Korea University, Seoul 02841, Republic of Korea

[b]School of Chemical and Biological Engineering, Institute of Chemical Processes, Seoul National University, Seoul 08826, Republic of Korea

[†]Equal contribution

Polymer Graph Convolutional Network architecture

In Figure 2 of main text, we depict the conversion process of a single bottlebrush copolymer into a feature matrix and an adjacency matrix. The nodes in the graph are represented by the beads obtained from the DPD simulation, while edges correspond to the bonds between these beads. The feature matrix encapsulates polymer information, including bead type and repulsion parameters. Specifically, the top three components of the feature matrix denote the bead types (eq S1). The subsequent components are allocated to represent repulsion parameters, ranging from 25 to 35 with an increment of 1 (eq S2).

$$v_{type\ A} : (1\ 0\ 0),\ v_{type\ B} : (0\ 1\ 0),\ v_{tyep\ C} : (0\ 0\ 1) \qquad \text{(eq S1)}$$

$$repulsion\ parameter\ (a_{ij}) = 26 : (0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0) \qquad \text{(eq S2)}$$

To organize the bead information systematically, numerical assignments are allocated to individual beads. In Figure S1, which serves as a simplified depiction of the bottlebrush copolymer, numerical labels are initially assigned to beads along the backbone chain. Subsequently, beads along the side chain are numbered sequentially form left to right.

Adjacency matrix contains bonding information between beads in the graph representation. In the adjacency matrix (A), if nodes i and j are connected, the corresponding component A_{ij} is assigned a value of 1. To preserve the individual information of each node, the diagonal elements of the adjacency matrix are set to 1, as depicted by the red elements in Figure 2 in main text. To match the feature matrix and the adjacency matrix, the matrix size is established based on the maximum value of short side chain length (N_{ss}). As a result, feature matrix size is fixed at 108×14 (108: The maximum number of beads of single bottlebrush

copolymer, 14: bead type and repulsion parameter information) and adjacency matrix size is fixed at 108×108.

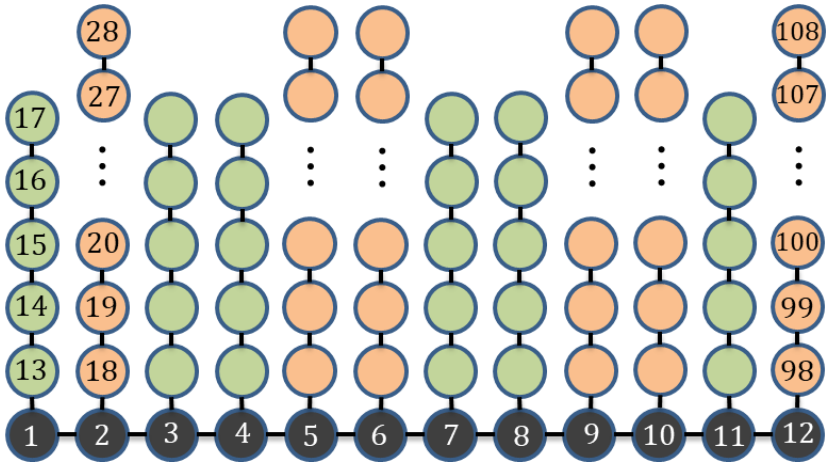

Figure S1. Numbering of beads for stacking feature vector

We provide a detailed process illustration for predicting single chain properties by using GCN method as shown in Figure S2. Initially, the resultant matrix, obtained by multiplying the feature matrix (F) and adjacency matrix (A), forms a distinctive representation of unique matrix sized 108×14 which encapsulates the architectural and repulsion parameter values inherent to the polymer (Figure S2a). Next, the unique matrix undergoes convolutional updates through weight and bias vectors. A weight matrix sized 14×64 and bias matrix sized 108×64 are employed to embed the unique matrix, resulting in a transformed embedded matrix H^0 sized 108×64 (Figure S2b). Then, embedded matrix H^0 is multiplied by the adjacency matrix to update information among connected nodes, and this is multiplied by weight matrix and bias matrix is added as described in the equation $H^{i+1} = \sigma(AH^iW^i + b^i)$. In this procedure, the graph is iteratively updated using weight matrix sized 64×64 and bias vectors sized 108×64, along with the adjacency matrix, as it passes through 5 hidden layers (Figure S2c). After convolution, a simple pooling is applied by summing along the columns, resulting in a transformed 1x64 matrix. Lastly, the matrix is converted to the target single chain physical property through linear regression using a 64x1 weight and bias vectors (Figure S2d).

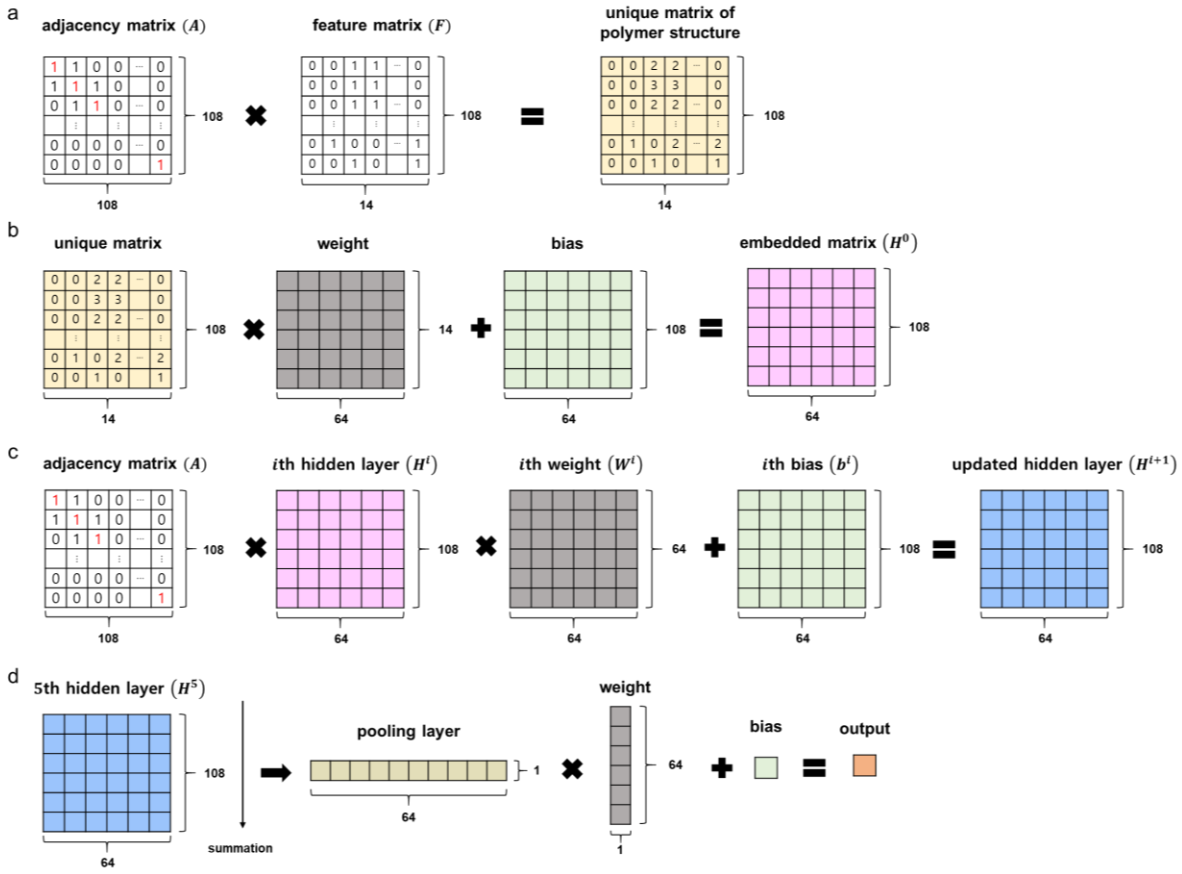

Figure S2. (a) Process of making the unique matrix for bottle brush copolymer. (b) Process of embedding unique polymer matrix. (c) Process of updating hidden layers. (d) Process of pooling and regression to predict the physical property (output).

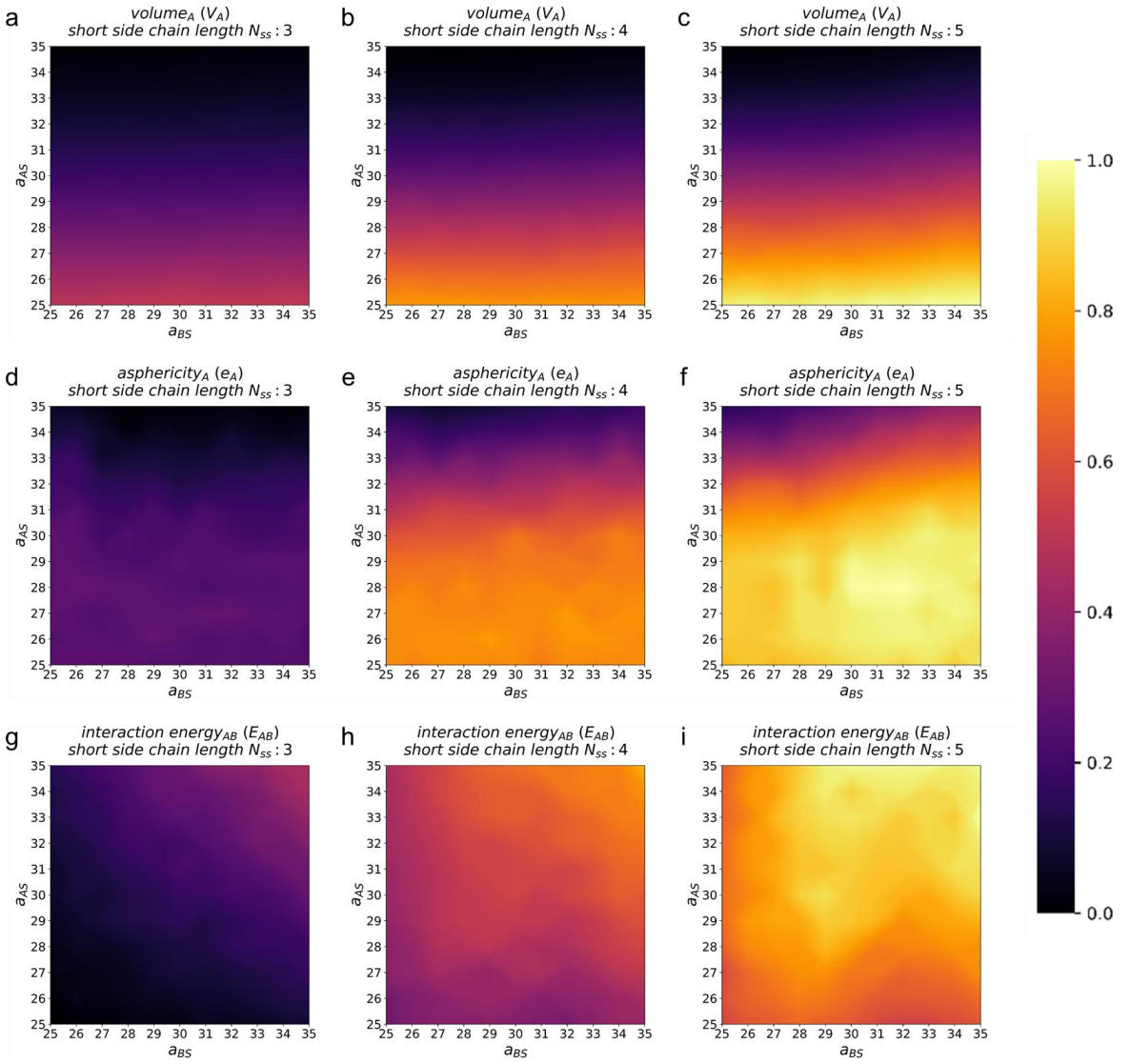

Figure S3. The normalized heatmap represents the convex hull volume of short side chain (V_A) of (a) short side chain length 3, (b) short side chain length 4, and (c) short side chain length 5. The normalized heatmap represents the asphericity of short side chain (e_A) of (d) short side chain length 3, (e) short side chain length 4, and (f) short side chain length 5. The normalized heatmap represents the interaction energy between short side chain-long side chain (E_{AB}) of (g) short side chain length 3, (h) short side chain length 4, and (i) short side chain length 5. The sequence is fixed at [001100110011].

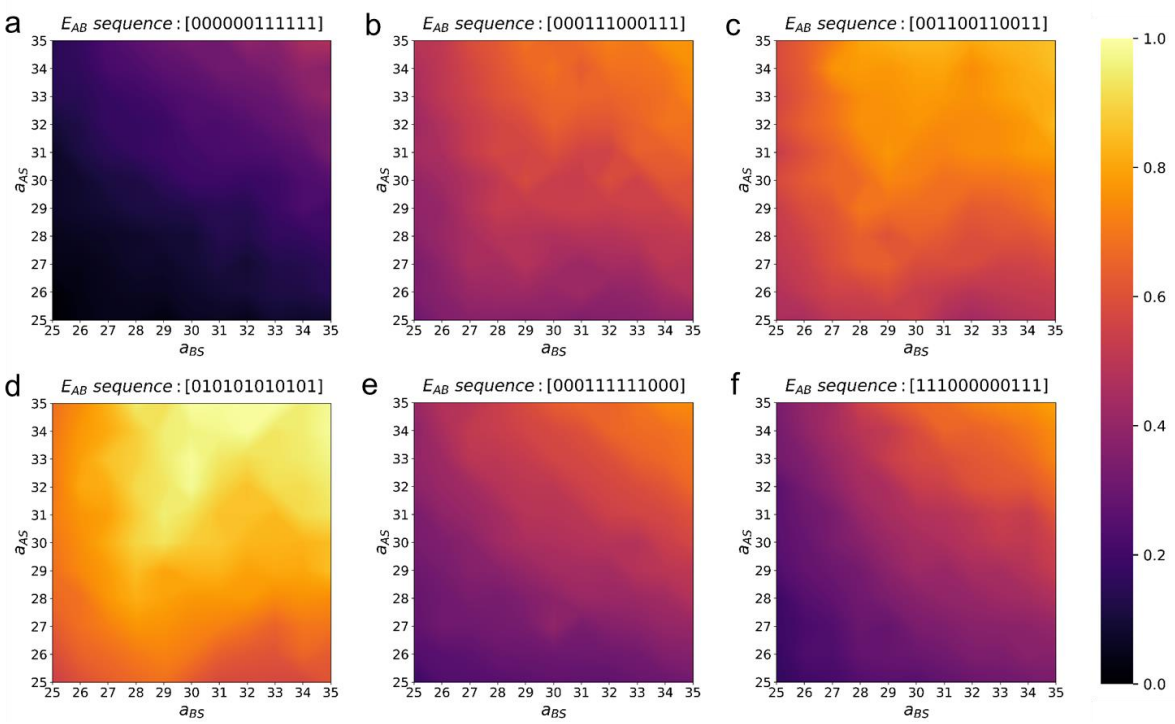

Figure S4. The normalized heatmap represents the interaction energy between short and long side chains (E_{AB}) of a single bottlebrush copolymer of short side chain length 4 with different sequences. (a) [000000111111]. (b) [000111000111]. (c) [001100110011]. (d) [010101010101]. (e) [000111111000]. (f) [111000000111].

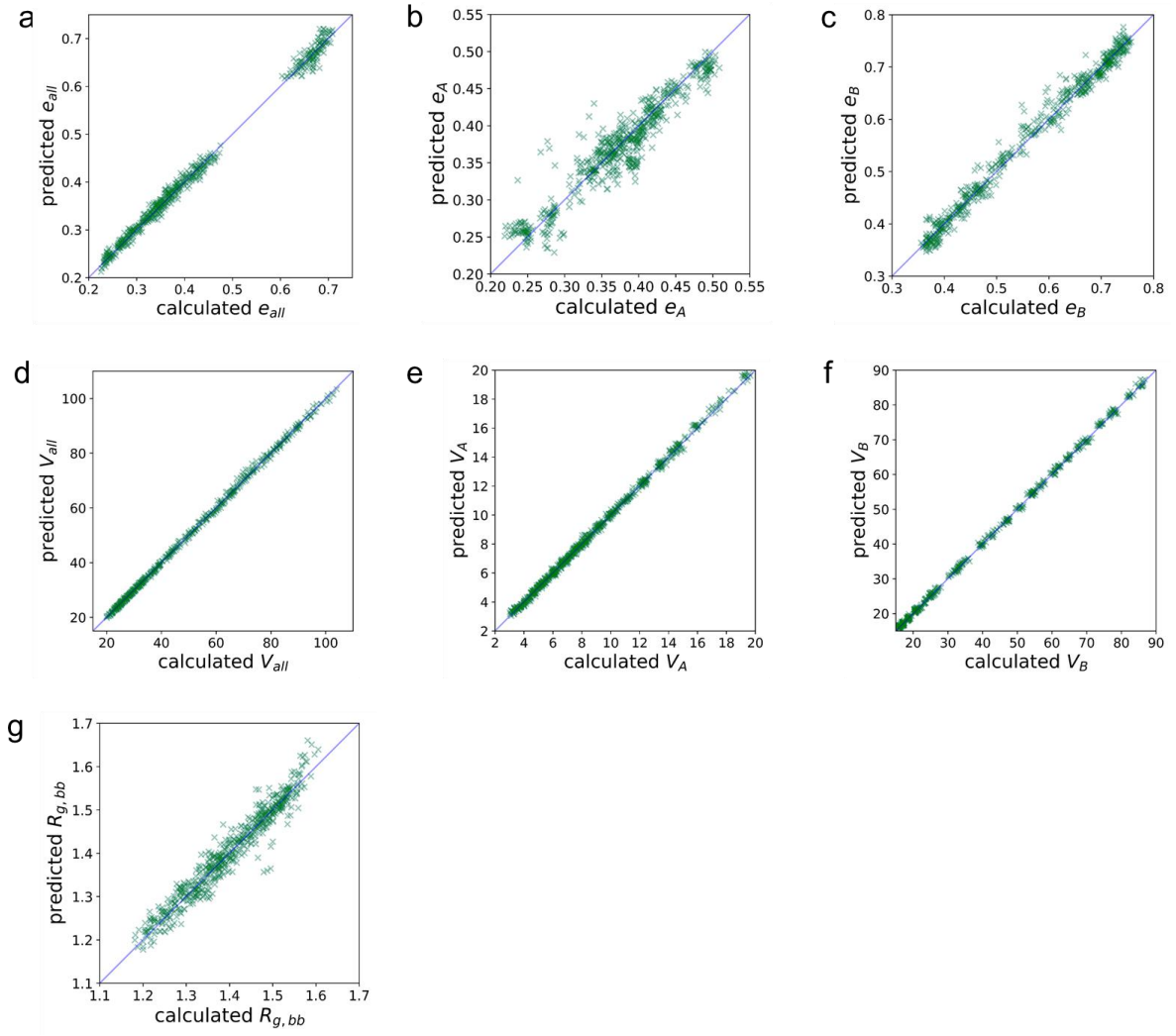

Figure S5. Predictive performance of 7 single chain physical properties; (a) Asphericity of all chain beads, (b) asphericity of short side chain (bead type A), (c) asphericity of long side chain (bead type B), (d) convex hull volume of all chain beads, (e) convex hull volume of short side chain (bead type A), (f) convex hull volume of long side chain (bead type B), and (g) radius of gyration of backbone chain (bead type C).

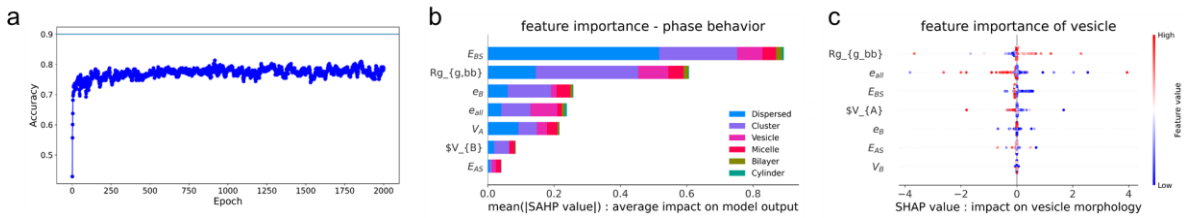

Figure S6. (a) The accuracy of classification with randomly chosen 7 input features. (b) The input feature importance for classification process. The color of each bar in the graph corresponds to the extent of its influence on specific self-assembled structure, matched by color, with longer bars indicating a greater impact on the classification process. (c) The feature importance for the formation of vesicle morphology. In the graph, red points indicate high values of physical properties, while blue points indicate low values, aiding in the classification process. If the value on the x-axis is positive, it signifies a high probability of predicting the formation of vesicles, whereas a negative value indicates a low probability of such prediction.

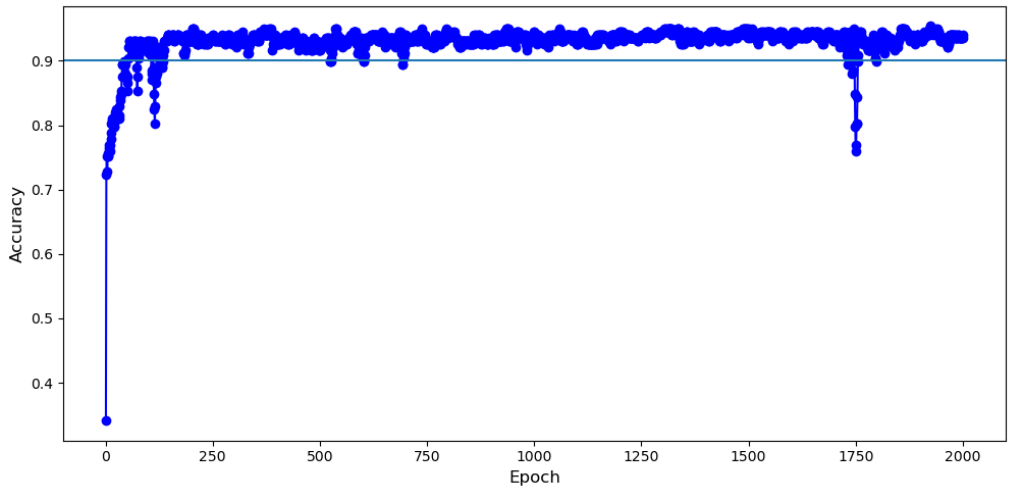

Figure S7. The morphology classification accuracy of validation set using GCN graphs as input features.

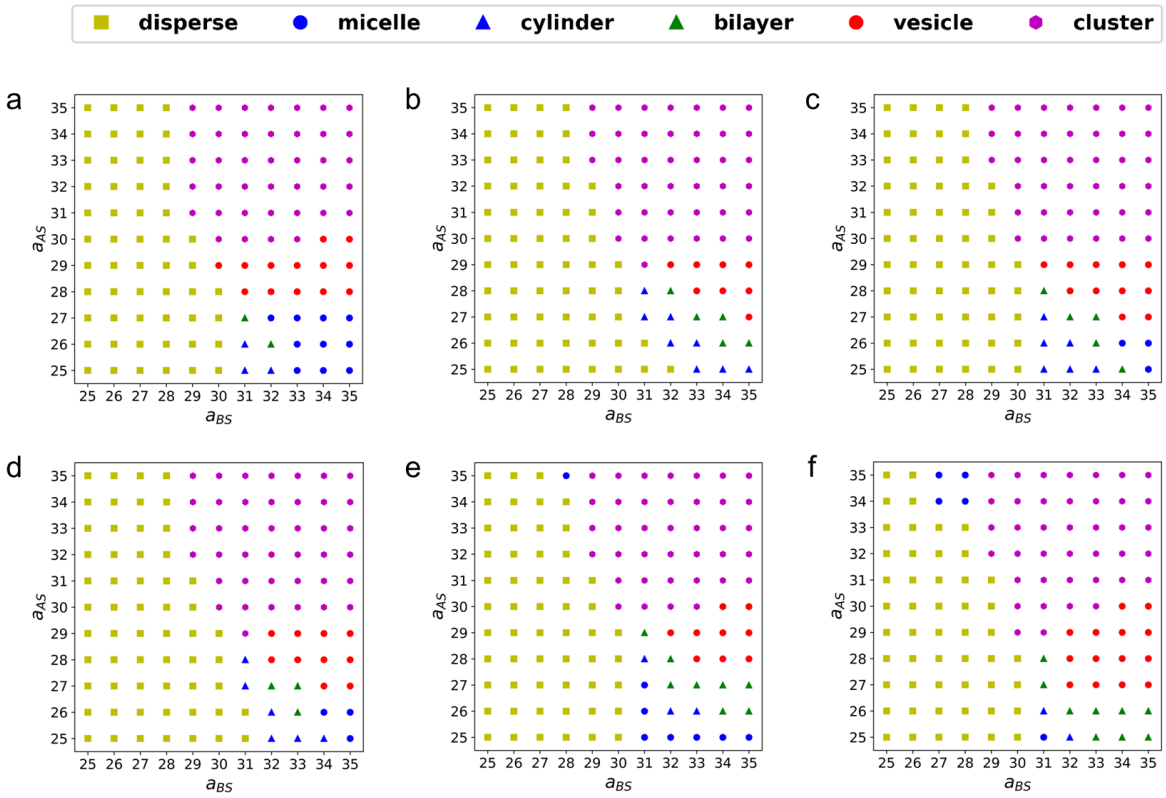

Figure S8. The phase diagram of bottlebrush copolymers with short side chain length 4 with different chain sequences from DPD simulation; (a) [000000111111], (b) [000111000111], (c) [001100110011], (d) [0101010101010], (e) [000111111000], and (f) [111000000111].

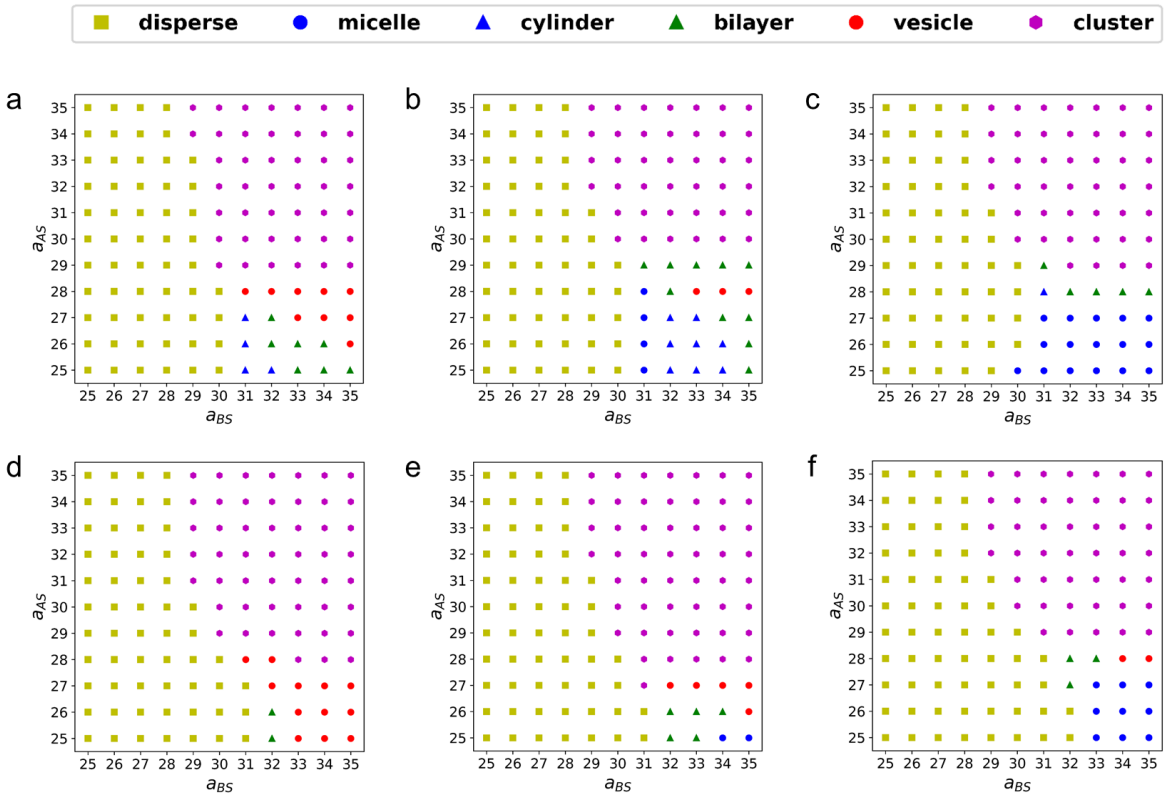

Figure S9. The upper phase diagrams (a-c) depict results observed through DPD simulation, with (a) short side chain length 3, (b) short side chain length 4, and (c) short side chain length 5 for the sequence [010100011101]. The lower phase diagrams (d-f) represent classification predictions using the predicted 7 physical properties as input data, with (d) short side chain length 3, (e) short side chain length 4, and (f) short side chain length 5 for the sequence [010100011101].